Issues in the Estimation of Mis-Specified Models of Fractionally Integrated Processes[†]

Gael M. Martin[‡], K. Nadarajah & D. S. Poskitt
Department of Econometrics & Business Statistics, Monash University

August 31, 2016

Abstract

In this paper we quantify the impact of model mis-specification on estimation in fractionally integrated processes. We show that four alternative parametric estimators – frequency domain maximum likelihood, Whittle, time domain maximum likelihood and conditional sum of squares – converge to the same pseudo-true value under common mis-specification, and that they will possess a common asymptotic distribution. The results are derived assuming appropriate closure of the parameter space for the estimated (mis-specified) fractional model, and under complete generality for the mis-specified model and the true data generating process, allowing for long memory, short memory and antipersistence in both the model and the process. As well as providing theoretical insights, we explore finite sample behaviour, with the initial set of simulation experiments conducted under the assumption of a known mean, as accords with the theoretical derivations. We demonstrate that the time-domain estimators have the smallest bias and mean squared error as estimators of the pseudo-true value of the long memory parameter, with conditional sum of squares being the most accurate estimator overall and having a relative efficiency that is approximately double that of frequency domain maximum likelihood, across a range of designs. The importance of the known mean assumption to this outcome is illustrated via the production of an alternative set of bias and MSE results, in which the estimators are applied to demeaned data.

Keywords and phrases: bias, conditional sum of squares, frequency domain, long memory models, maximum likelihood, mean squared error, pseudo true parameter, time domain, Whittle.

MSC2010 subject classifications: Primary 62M10, 62M15; Secondary 62G09
JEL classifications: C18, C22, C52

1 Introduction

Let y_t, $t \in \mathbb{Z}$, be a stationary Gaussian process with mean μ_0 and spectral density $f_0(\lambda)$, $\lambda \in [-\pi, \pi]$, that is such that

$$f_0(\lambda) \sim |\lambda|^{-2d_0} L_0(\lambda) \quad \text{as} \quad \lambda \to 0\,,$$

where $0 < |d_0| < 0.5$ and $L_0(\lambda)$ is a positive function that is slowly varying at 0. Prototypical examples of processes of this type are fractional Gaussian noise, obtained as the increments of

[†]This research has been supported by Australian Research Council (ARC) Discovery Grants DP120102344 and DP150101728. The authors would like to thank the Editor and two anonymous referees for very helpful and constructive comments on an earlier draft of the paper.

[‡]Corresponding author: Gael Martin, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia. Tel.: +61-3-9905-1189; fax: +61-3-9905-5474; email: gael.martin@monash.edu.

self-similar processes, and fractional autoregressive moving average processes. The process y_t is said to exhibit long memory (or long-range dependence) when $0 < d_0 < 0.5$, short memory (or short-range dependence) when $d_0 = 0$, and antipersistence when $-0.5 < d_0 < 0$, and in this paper we undertake an examination of the consequences of model mis-specification when modelling such processes.

We begin by establishing that four alternative parametric techniques – frequency domain maximum likelihood (FML), Whittle, time domain maximum likelihood (TML) and conditional sum of squares (CSS) – when used to estimate the parameters of a fractionally integrated model in which the short memory component is mis-specified, converge to a common pseudo-true parameter value. Convergence is established for all three forms of dependence in the TDGP - long memory, short memory and antipersistence - and for appropriate closure of the parameter space of the fitted (mis-specified) model. We then establish that all four estimators converge in distribution to a common asymptotic distribution under mis-specification. This complements earlier work in which the properties of these estimators, including their asymptotic equivalence, is established for *correctly specified* long memory models, see Fox and Taqqu (1986), Dahlhaus (1989), Sowell (1992), Beran (1995) and Robinson (2006), among others. It also represents a significant extension of the analysis in Yajima (1992) and Chen and Deo (2006), wherein the properties of the TML estimator of autoregressive moving average (ARMA) models applied to fractional noise, and the FML estimator of mis-specified models of long memory processes, are examined respectively. Our work also complements that of Cavaliere, Nielsen and Taylor (2015), in which fractional models with correctly specified transfer functions but mis-specified variance structures are the focus.

First, we show that the objective functions that define all four estimators are divergent in a subset of the parameter space, and that otherwise, under common mis-specification, these functions converge to simple functions of a common limit. The estimators are, accordingly, shown to converge to the same pseudo-true parameter value – by definition the value that optimizes the limiting objective function. Secondly, we derive closed-form representations for the first-order conditions that define the pseudo-true parameters for general autoregressive fractionally integrated moving average (ARFIMA) model structures – both true and mis-specified. This extends the analysis in Chen and Deo (2006), in which expressions for the first-order conditions that define the pseudo-true parameter are provided for the FML estimator only, and for certain special cases. Thirdly, we extend the asymptotic theory established by these authors for the FML estimator to the other three estimators and show that all four methods are asymptotically equivalent, in that they will converge in distribution under common mis-specification. A fourth contribution is the provision of a closed-form representation of the (common) asymptotic distribution that obtains under the most extreme type of mis-specification – whereby both a $\sqrt{n}$ rate of convergence and limiting Gaussianity is lost – together with a demonstration of how to implement the distribution numerically using appropriate truncation of the series expansion that characterizes the distribution. This then enables us to illustrate graphically the differences in the rates at which the finite sample distributions of the four different estimators approach the (common) asymptotic distribution. Notably, when the difference between the true and pseudo-true values of d is greater than or equal to 0.25, there is a distinct grouping into frequency domain and time-domain techniques; with the latter tending to replicate the asymptotic distribution more closely than the former in small samples. Finally, we perform an extensive set of simulation experiments in which the relative finite sample performance of all four mis-specified estimators is assessed. The experiments are first conducted assuming a known (zero) mean, in line with the theoretical derivations in the paper, and then re-run with the mean estimated. The ranking of the estimators, in terms of bias and mean squared error, is shown to depend heavily on whether the mean is specified or estimated, a conclusion that parallels results documented previously for correctly specified ARFIMA models (see, for example Cheung and Diebold, 1994; Nielsen and Frederiksen, 2005; Sowell, 1992).

In defining the Whittle estimator we focus on a particular approximation to the frequency domain Gaussian (negative) log likelihood, in which *sums* over Fourier frequencies are used to approximate the relevant integrals. Despite the analytical equivalence of this estimator with the FML estimator for large n, the small sample performances of the two procedures will be seen to differ systematically. In common with the FML approach, this form of Whittle estimator is invariant to the mean of the process. For interest, we also present selected numerical results pertaining to the integral-based form of the Whittle estimator (referred to hereafter as 'exact Whittle'), both for the known mean case and when the mean is unknown, the lack of invariance of this estimator to the mean rendering this latter exercise of particular interest.

The paper is organized as follows. In Section 2 we define the estimation problem, namely producing an estimate of the parameters of a fractionally integrated model when the component of the model that characterizes the short term dynamics is mis-specified. The criterion functions that define the Whittle, TML and CSS estimators, as well as the FML estimator, are specified, and we demonstrate that all four estimators possess a common probability limit under mis-specification. The limiting form of the criterion function for a mis-specified ARFIMA model is presented in Section 3, under complete generality for the short memory dynamics in the true process and estimated model, and closed-form expressions for the first-order conditions that define the pseudo-true values of the parameters are then given. The asymptotic equivalence of all four estimation methods is proved in Section 4. The finite sample performance of the alternative estimators of d in the mis-specified model – with reference to estimating the pseudo-true value of d – is documented in Section 5. The form of the sampling distribution is recorded, as is the bias and mean squared error (MSE), under different degrees of mis-specification, for all four estimators. Bias and MSE results are also documented for the exact Whittle estimator. The experiments are first conducted assuming a known (zero) mean, in line with the theoretical derivations in the paper. In this case, the CSS estimator exhibits superior performance, in terms of bias and mean squared error, across a range of mis-specification settings, whilst the performance of the FML estimator is notably inferior. We then re-run the simulations using demeaned data. Only the time domain estimators, plus the exact Whittle estimator, are affected by this change, and with the rate of convergence of the sample mean being slow under long memory we find that the superiority of the time domain estimators is diminished - the (sums-based) Whittle estimator now being the best performer overall. The paper concludes in Section 6 with a brief summary and some discussion of several issues that arise from the work. The proofs of the results presented in the paper are assembled in Appendix A. Appendix B contains certain technical derivations referenced in the text.

2 Estimation Under Mis-specification

Assume that $\{y_t\}$ is generated from a TDGP that is a stationary Gaussian process with spectral density given by

$$\frac{\sigma_0^2}{2\pi} f_0(\lambda) = \frac{\sigma_0^2}{2\pi} g_0(\lambda) \left(2\sin(\lambda/2)\right)^{-2d_0}, \tag{1}$$

where σ_0^2 is the innovation variance, $g_0(\lambda)$ is a real valued symmetric function of λ defined on $[-\pi, \pi]$ that is bounded above and bounded away from zero, and $-0.5 < d_0 < 0.5$. The model refers to a parametric specification for the spectral density of $\{y_t\}$ of the form

$$\frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda) = \frac{\sigma^2}{2\pi} g_1(\boldsymbol{\beta}, \lambda) \left(2\sin(\lambda/2)\right)^{-2d}, \tag{2}$$

where $g_1(\boldsymbol{\beta}, \lambda)$ is a real valued symmetric function of λ defined on $[-\pi, \pi]$. The parameter of interest will be taken as $\boldsymbol{\eta} = (d, \boldsymbol{\beta}^T)^T$, where $d \in (-0.5, 0.5)$ and $\boldsymbol{\beta} \in \mathbb{B}$, where $\mathbb{B}$ is an l-dimensional bounded, connected set in $\mathbb{R}^l$. The variance σ^2 will be viewed as a supplementary

or nuisance parameter. The model is to be estimated from a realization y_t, $t = 1, \ldots, n$, of $\{y_t\}$ and it will be assumed that:

$(A.1)$ $g_1(\boldsymbol{\beta},\lambda)$ is thrice differentiable with continuous third derivatives.

$(A.2)$ $\inf_{\boldsymbol{\beta}} \inf_{\lambda} g_1(\boldsymbol{\beta},\lambda) > 0$ and $\sup_{\boldsymbol{\beta}} \sup_{\lambda} g_1(\boldsymbol{\beta},\lambda) < \infty$.

$(A.3)$ $\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left| \frac{\partial g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i} \right| < \infty$, $1 \leqslant i \leqslant l$.

$(A.4)$ $\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left| \frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \beta_j} \right| < \infty$, $\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left| \frac{\partial^2 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \lambda} \right| < \infty$, $1 \leqslant i, j \leqslant l$.

$(A.5)$ $\sup_{\lambda} \sup_{\boldsymbol{\beta}} \left| \frac{\partial^3 g_1(\boldsymbol{\beta},\lambda)}{\partial \beta_i \partial \beta_j \partial \beta_k} \right| < \infty$, $1 \leqslant i, j, k \leqslant l$.

$(A.6)$ $\int_{-\pi}^{\pi} \log g_1(\boldsymbol{\beta},\lambda) d\lambda = 0$ for all $\boldsymbol{\beta} \in \overline{\mathbb{B}}$ where $\overline{\mathbb{B}}$ denotes the closure of $\mathbb{B}$.

The above data structure and model characterization, and Assmptions A.1-A.6, correspond to those adopted in Chen and Deo (2006), save that here the fractional indices are not constrained to be non-negative. If there exists a subset of $[-\pi, \pi]$ with non-zero Lebesgue measure in which $g_1(\boldsymbol{\beta},\lambda) \neq g_0(\lambda)$ for all $\boldsymbol{\beta} \in \overline{\mathbb{B}}$ then the model will be referred to as a mis-specified model (MisM).

An ARFIMA model for a time series $\{y_t\}$ may be defined as follows,

$$\phi(L)(1-L)^d\{y_t - \mu\} = \theta(L)\varepsilon_t, \tag{3}$$

where $\mu = E(y_t)$, L is the lag operator such that $L^k y_t = y_{t-k}$, and $\phi(z) = 1 + \phi_1 z + \ldots + \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$ are the autoregressive and moving average operators respectively, where it is assumed that $\phi(z)$ and $\theta(z)$ have no common roots and that the roots lie outside the unit circle. The errors $\{\varepsilon_t\}$ are assumed to be a white noise sequence with finite variance $\sigma_\varepsilon^2 > 0$. For $|d| < 0.5$, $\{y_t\}$ can be represented as an infinite-order moving average of $\{\varepsilon_t\}$ with square-summable coefficients and, hence, on the assumption that the specification in (3) is correct, $\{y_t\}$ is defined as the limit in mean square of a covariance-stationary process. We will therefore assume that $0 < |d| < 0.5$. When $0 < d < 0.5$ neither the moving average coefficients nor the autocovariances of the process are absolutely summable, declining at a slow hyperbolic rate rather than the exponential rate typical of an ARMA process, with the term 'long memory' invoked accordingly. A detailed outline of the properties of such ARFIMA processes is provided in Beran (1994). For an ARFIMA model we have $g_1(\boldsymbol{\beta}, \lambda) = |\theta(e^{i\lambda})|^2 / |\phi(e^{i\lambda})|^2$ where $\boldsymbol{\beta} = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q)^T$ and an $ARFIMA(p, d, q)$ model will be mis-specified if the realizations are generated from a true $ARFIMA(p_0, d_0, q_0)$ process and any of $\{p \neq p_0 \cup q \neq q_0\} \setminus \{p_0 \leq p \cap q_0 \leq q\}$ obtain.

The estimators to be considered are all to be obtained by minimizing an objective function, $Q_n(\boldsymbol{\eta})$ say, and under mis-specification the estimator, denoted by $\widehat{\boldsymbol{\eta}}_1$, is obtained by minimizing $Q_n(\boldsymbol{\eta})$ on the assumption that $\{y_t\}$ follows the MisM.[1] For any given $Q_n(\boldsymbol{\eta})$ there exists a non-stochastic limiting objective function $Q(\boldsymbol{\eta})$ that is independent of the sample size such that $|Q_n(\boldsymbol{\eta}) - Q(\boldsymbol{\eta})| \to^p 0$, and provided certain conditions hold, $Q_n(\widehat{\boldsymbol{\eta}}_1)$ will converge to $Q(\boldsymbol{\eta}_1)$ where $\boldsymbol{\eta}_1$ is the minimizer of $Q(\boldsymbol{\eta})$, and $\widehat{\boldsymbol{\eta}}_1 \to^p \boldsymbol{\eta}_1$ as a consequence. In Subsection 2.1 we specify the form of $Q_n(\boldsymbol{\eta})$ associated with the FML estimator considered

[1] We follow the usual convention by denoting the estimator obtained under mis-specification as $\widehat{\boldsymbol{\eta}}_1$ rather than simply by $\widehat{\boldsymbol{\eta}}$, say. This is to make it explicit that the estimator is obtained under mis-specification and does not correspond to the estimator produced under the correct specification of the model, which could be denoted by $\widehat{\boldsymbol{\eta}}_0$.

in Chen and Deo (2006) and outline its relationship with the criterion functions underlying two alternative versions of the frequency domain estimator introduced by Whittle, making it clear which form of Whittle estimator is the focus of our theoretical investigations. In Subsection 2.2 we define the two time domain estimators that we consider here, TML and CSS. In Subsection 2.3 we derive our asymptotic results pertaining to the convergence of $Q_n(\boldsymbol{\eta})$ to $Q(\boldsymbol{\eta})$ and demonstrate the equivalence of the limiting criterion functions of the alternative estimators to the limiting criterion function of the FML estimator. Hence, the value that minimizes the limiting criterion function of all four estimators is shown to be identical, and the asymptotic convergence of $\widehat{\boldsymbol{\eta}}_1$ to $\boldsymbol{\eta}_1$ for the four estimators is established. In the theoretical derivations we adopt the assumption of a known mean for both the true and estimated models, with a zero value specified without loss of generality.

2.1 Frequency domain estimators

In their paper Chen and Deo (2006) focus on the estimator of $\boldsymbol{\eta} = (d, \boldsymbol{\beta}^T)^T$ defined as the value of $\boldsymbol{\eta}$, $\widehat{\boldsymbol{\eta}}_1$, that minimizes the objective function

$$Q_n^{(1)}(\boldsymbol{\eta}) = \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\,, \tag{4}$$

where $I(\lambda_j)$ is the periodogram, defined as $I(\lambda) = \frac{1}{2\pi n}|\sum_{t=1}^{n} y_t \exp(-i\lambda t)|^2$ evaluated at the Fourier frequencies $\lambda_j = 2\pi j/n$; $(j = 1, ..., \lfloor n/2 \rfloor)$, $\lfloor x \rfloor$ is the largest integer not greater than x. We have labeled this the FML estimator. The objective function in (4) is an approximation to the frequency domain Gaussian (negative) log-likelihood introduced initially by Whittle (1953) for short-range dependent processes, namely

$$W_n(\sigma^2, \boldsymbol{\eta}) = \int_{-\pi}^{\pi} \left\{ \log \frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda) + \frac{2\pi I(\lambda)}{\sigma^2 f_1(\boldsymbol{\eta}, \lambda)} \right\} d\lambda\,, \tag{5}$$

and it coincides with the frequency domain objective function considered in Hannan (1973). Concentrating out σ^2 in (5) and minimizing the associated profile function with respect to $\boldsymbol{\eta}$ produces what we refer to as the exact Whittle estimator.

An alternative approximation to the Whittle criterion function in (5), considered for example in Beran (1994), is

$$Q_n^{(2)}(\sigma^2, \boldsymbol{\eta}) = \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log \left[\frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta},\lambda_j) \right] + \frac{(2\pi)^2}{\sigma^2 n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}. \tag{6}$$

Taking $\boldsymbol{\eta}$ as the parameter of interest and concentrating $Q_n^{(2)}(\sigma^2, \boldsymbol{\eta})$ with respect to σ^2 indicates that the value of σ^2 that minimises (6) is given by $\widehat{\sigma}^2(\boldsymbol{\eta}) = 2Q_n^{(1)}(\boldsymbol{\eta})$. Substituting back in to (6) yields the (negative) profile likelihood,

$$Q_n^{(2)}(\boldsymbol{\eta}) = \frac{2\pi}{2} \log \left(\frac{\widehat{\sigma}^2(\boldsymbol{\eta})}{2\pi} \right) + \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) + \pi.$$

Minimization of $Q_n^{(2)}(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ yields what we call (simply) the Whittle estimator, and which is the form of Whittle procedure that features in our theoretical derivations. Since $\lim_{n\to\infty} \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \log f_1(\boldsymbol{\eta}, \lambda_j) = 0$ (see Appendix A.3.1) it follows that this estimator is equivalent to the FML estimator for large n. However, as indicated in Boes, Davis and Gupta (1989), and as will be seen in the simulation results documented in Section 5, the finite sample performance of these two estimators differs. For interest we also report selected numerical results on the finite sample performance of the exact Whittle estimator described above.

2.2 Time domain estimators

The criterion functions of the two alternative time domain estimators are defined as follows:

- Let $\mathbf{Y}^T = (y_1, y_2, ..., y_n)$ and denote the variance covariance matrix of $\mathbf{Y}$ derived from the mis-specified model by $\sigma^2 \mathbf{\Sigma}_\eta = [\gamma_1 (i-j)]$, $i, j = 1, 2, ..., n$, where

$$\gamma_1(\tau) = \gamma_1(-\tau) = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} f_1(\boldsymbol{\eta}, \lambda) e^{i\lambda\tau} d\lambda \,.$$

  The Gaussian log-likelihood function for the TML estimator is

$$-\frac{1}{2}\left(n \log(2\pi\sigma^2) + \log|\mathbf{\Sigma}_\eta| + \frac{1}{\sigma^2} (\mathbf{Y} - \mu\mathbf{l})^T \mathbf{\Sigma}_\eta^{-1} (\mathbf{Y} - \mu\mathbf{l}) \right) , \tag{7}$$

  where $\mathbf{l} = (1, 1, ..., 1)$, and maximizing (7) is equivalent to minimizing the criterion function

$$Q_n^{(3)}(\sigma^2, \boldsymbol{\eta}) = \log\sigma^2 + \frac{1}{n}\log|\mathbf{\Sigma}_\eta| + \frac{1}{n\sigma^2} (\mathbf{Y} - \mu\mathbf{l})^T \mathbf{\Sigma}_\eta^{-1} (\mathbf{Y} - \mu\mathbf{l}) \,. \tag{8}$$

- To construct the CSS estimator note that we can expand $(1-z)^d$ in a binomial expansion as

$$(1-z)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} z^j \,, \tag{9}$$

  where $\Gamma(\cdot)$ is the gamma function. Furthermore, since $g_1(\boldsymbol{\beta}, \lambda)$ is bounded, by Assumption $(A.2)$, we can employ the method of Whittle (Whittle, 1984, §2.8) to construct an autoregressive operator $\alpha(\boldsymbol{\beta}, z) = \sum_{i=0}^{\infty} \alpha_i(\boldsymbol{\beta}) z^i$ such that $g_1(\boldsymbol{\beta}, \lambda) = |\alpha(\boldsymbol{\beta}, e^{i\lambda})|^{-2}$. The objective function of the CSS estimation method then becomes

$$Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n} \sum_{t=1}^{n} e_t^2 \,, \tag{10}$$

  where

$$e_t = \sum_{i=0}^{t-1} \tau_i(\boldsymbol{\eta}) (y_{t-i} - \mu) \tag{11}$$

  and the coefficients $\tau_j(\boldsymbol{\eta})$, $j = 0, 1, 2, \ldots$, are given by $\tau_0(\boldsymbol{\eta}) = 1$ and

$$\tau_j(\boldsymbol{\eta}) = \sum_{s=0}^{j} \frac{\alpha_{j-s}(\boldsymbol{\beta})\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \,, \quad j = 1, 2, \ldots \,. \tag{12}$$

  As with the FML estimator, the CSS estimate of σ^2 is given implicitly by the minimum value of the criterion function.

We can think of the CSS estimator as providing an approximation to the TML estimator that parallels the approximation of the FML and (sums-based) Whittle estimators to the exact Whittle estimator.

2.3 Convergence Properties

In Chen and Deo (2006) it is shown that if $\{y_t\}$ is a long-range dependent process, then on subsets of the parameter space of the form $(\delta, 0.5-\delta) \times \Phi$, where $0 < \delta < 0.25$ and Φ is a compact convex set, we have $\text{plim}_{n\to\infty} |Q_n^{(1)}(\boldsymbol{\eta}) - Q(\boldsymbol{\eta})| = 0$, where

$$Q(\boldsymbol{\eta}) = \lim_{n\to\infty} E_0 \left[Q_n^{(1)}(\boldsymbol{\eta}) \right] = \frac{\sigma_0^2}{2\pi} \int_0^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda \tag{13}$$

under Assumptions $A.1 - A.3$ (Chen and Deo, 2006, Lemma 2). Here, and in what follows, the zero subscript denotes that the moments are defined with respect to the TDGP. The limiting objective function $Q(\boldsymbol{\eta})$ then defines a pseudo-true parameter value to which the FML estimator will converge, since with the addition of the assumption that there exists a unique vector $\boldsymbol{\eta}_1 = (d_1, \boldsymbol{\beta}_1^T)^T \in (\delta, 0.5-\delta) \times \Phi$ that minimizes $Q(\boldsymbol{\eta})$, it follows that the FML estimator will converge to $\boldsymbol{\eta}_1$.

Here we will establish that on the intersection of the closure of $\mathbb{D} \times \mathbb{B}$ with the set $\{\boldsymbol{\eta} : (d_0 - d) < 0.5\}$ where $\mathbb{D} \subseteq (-0.5, 0.5)$, we have $\lim_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) = Q(\boldsymbol{\eta})$ almost surely $(a.s.)$ where $Q(\boldsymbol{\eta})$ is defined as in (13). This is the content of Lemmas 1 and 2 and Proposition 1 below. Proposition 2 establishes that the FML estimator converges to $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$. We will also establish that the probability limit of both $Q_n^{(2)}(\sigma^2, \boldsymbol{\eta})$ and $Q_n^{(3)}(\sigma^2, \boldsymbol{\eta})$ is $\log \sigma^2 + 2Q(\boldsymbol{\eta})/\sigma^2$, and that $\text{plim}_{n\to\infty} = Q_n^{(4)}(\boldsymbol{\eta}) = 2Q(\boldsymbol{\eta})$, and, hence, that the probability limit of each of the three alternative estimators, Whittle, TML and CSS, is equivalent to that of the FML estimator. This is the content of Theorem 1.

Index by $i = 1, 2, 3$ and 4 the estimators associated with the FML, Whittle, TML and CSS criterion functions respectively; that is $\widehat{\boldsymbol{\eta}}_1^{(i)}$ minimises $Q_n^{(i)}(\cdot)$, $i = 1, 2, 3, 4$, with each viewed as a function of $\boldsymbol{\eta}$. Suppose also that the following assumption holds.

$(A.7)$ Let $\mathbb{E} = \overline{\mathbb{D} \times \mathbb{B}}$ denote the closure of $\mathbb{D} \times \mathbb{B}$. There exists a unique vector $\boldsymbol{\eta}_1 = (d_1, \boldsymbol{\beta}_1^T)^T \in \mathbb{E}$ which satisfies $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$.

Then with $\boldsymbol{\eta}$ being the parameter of interest, we can state the following theorem:

Theorem 1 *Suppose that the TDGP of $\{y_t\}$ is a Gaussian process with a spectral density as given in* (1) *and that the MisM in* (2) *satisfies Assumptions $A.1 - A.3$, $A.6$ and $A.7$. Let $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 1, 2, 3, 4$, denote, respectively, the FML, Whittle, TML and CSS estimators of the parameter vector $\boldsymbol{\eta} = (d, \boldsymbol{\beta}^T)^T$ of the MisM. Then $plim_{n\to\infty} \|\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(j)}\| = 0$ for all $i, j = 1, 2, 3, 4$, where the common limiting value of $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 1, 2, 3, 4$, is $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$.*

The proof of Theorem 1 relies upon the following two lemmas and their related propositions.

Lemma 1 *Suppose that the TDGP of $\{y_t\}$ is a Gaussian process with a spectral density as given in* (1) *and that the MisM in* (2) *satisfies Assumptions $A.1 - A.3$. Then*

(i) If $(d_0 - d) \geq 0.5$ for all $\boldsymbol{\eta} \in \mathbb{E}$ then $Q_n^{(1)}(\boldsymbol{\eta}) = O(n^{2(d_0-d)-1})$ uniformly in $\boldsymbol{\eta} \in \mathbb{E}$.

(ii) If $(d_0 - d) < 0.5$ for all $\boldsymbol{\eta} \in \mathbb{E}$ then for all $\delta > 0$

$$\left| \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j) + \delta} - \frac{\sigma_0^2}{2\pi} \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda) + \delta} d\lambda \right|$$

converges to zero almost surely and uniformly in $\boldsymbol{\eta} \in \mathbb{E}$.

Since, obviously, $f_1(\boldsymbol{\eta}, \lambda) < f_1(\boldsymbol{\eta}, \lambda) + \delta$ it follows from part (ii) of Lemma 1 that,

$$\begin{aligned} \liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) &\geq \lim_{n\to\infty} \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j) + \delta} \\ &= \frac{\sigma_0^2}{2\pi} \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda) + \delta} \quad a.s. \end{aligned}$$

uniformly in $\boldsymbol{\eta} \in \mathbb{E}$ when $(d_0 - d) < 0.5$. Letting $\delta \to 0$ and applying Lebegue's monotone convergence theorem gives

$$\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}) \geq Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi} \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda \quad a.s.$$

To establish that $Q(\boldsymbol{\eta})$ also provides a limit superior for $Q_n^{(1)}(\boldsymbol{\eta})$ when $(d_0 - d) < 0.5$, we will use the following lemma.

Lemma 2 *Suppose that the conditions of Lemma 1 hold and assume that* $(d_0 - d) < 0.5$ *for all* $\boldsymbol{\eta} \in \mathbb{E}$. *Set*

$$h_1(\boldsymbol{\eta}, \lambda) = \begin{cases} f_1(\boldsymbol{\eta}, \lambda), & f_1(\boldsymbol{\eta}, \lambda) \geq \delta \\ \delta, & f_1(\boldsymbol{\eta}, \lambda) < \delta\,. \end{cases}$$

Then for all $\delta > 0$

$$\left| \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} - \frac{\sigma_0^2}{2\pi} \int_0^{\pi} \frac{f_0(\lambda)}{h_1(\boldsymbol{\eta}, \lambda)} d\lambda \right|$$

converges to zero almost surely uniformly in $\boldsymbol{\eta} \in \mathbb{E}$.

Observe that $f_1(\boldsymbol{\eta}, \lambda) > 0$ when $d \geq 0$ and hence for δ sufficiently small we have $h_1(\boldsymbol{\eta}, \lambda) = f_1(\boldsymbol{\eta}, \lambda)$ for all $\lambda \in [-\pi, \pi]$. It follows immediately from Lemma 2 that

$$\lim_{n \to \infty} Q_n^{(1)}(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi} \int_0^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda = Q(\boldsymbol{\eta}) \quad a.s.$$

uniformly in $\eta \in E$ when $(d_0 - d) < 0.5$ and $d \geq 0$. We have thus established the following proposition, Proposition 1, in the case where $(d_0 - d) < 0.5$ and $d \geq 0$, (*cf.* Chen and Deo, 2006, Lemma 2). That this property also holds when $(d_0 - d) < 0.5$ and $d < 0$ is established in the proof of Proposition 1 presented in Appendix A.2.1. As such, the limiting form of the FML criterion function, given initially by Chen and Deo (2006) for the case $d, d_0 > 0$ only, is now shown to hold much more generally and, hence, to encompass all three forms of memory - long memory, short memory and antipersistence - in both the true and estimated models.

Proposition 1 *Suppose that the TDGP of* $\{y_t\}$ *is a Gaussian process with a spectral density as given in* (1) *and that the MisM in* (2) *satisfies Assumptions* $A.1 - A.3$. *Then if* $(d_0 - d) < 0.5$ *for all* $\boldsymbol{\eta} \in \mathbb{E}$,

$$\lim_{n \to \infty} Q_n^{(1)}(\boldsymbol{\eta}) = Q(\boldsymbol{\eta})$$

almost surely and uniformly in $\boldsymbol{\eta} \in \mathbb{E}$.

The next proposition follows as an almost immediate corollary of the previous developments (*cf.* Chen and Deo, 2006, Corollary 1) and establishes the convergence of the FML estimator to $\boldsymbol{\eta}_1$ under the same generality for both the true and mis-specified models as highlighted above.

Proposition 2 *Suppose that the TDGP of* $\{y_t\}$ *is a Gaussian process with a spectral density as given in* (1) *and that the MisM in* (2) *satisfies Assumptions* $A.1 - A.3$ *and* $A.7$. *Then if* $(d_0 - d) < 0.5$ *for all* $\boldsymbol{\eta} \in \mathbb{E}$, $\lim_{n \to \infty} Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)}) = Q(\boldsymbol{\eta}_1)$ *and* $\widehat{\boldsymbol{\eta}}_1^{(1)} \to \boldsymbol{\eta}_1$ *almost surely.*

Finally, as is clear from part (i) of Lemma 1, when $(d_0 - d) \geq 0.5$, $Q_n^{(1)}(\boldsymbol{\eta})$ diverges and, hence, there exists no pseudo-true parameter value $\boldsymbol{\eta}_1 = (d_1, \boldsymbol{\beta}_1^T)^T$ with $(d_0 - d_1) \geq 0.5$ to which $\widehat{\boldsymbol{\eta}}_1^{(1)}$ will converge. Given the relationship between $Q_n^{(1)}(\boldsymbol{\eta})$ and $Q_n^{(i)}(\cdot)$, $i = 2, 3, 4$ it follows that the same statement can also be made about the convergence behaviour of $\widehat{\boldsymbol{\eta}}_1^{(i)}$, $i = 2, 3, 4$ in this part of the parameter space.

Note that if the MisM were used to construct a one-step-ahead prediction, the mean squared prediction error would be

$$2Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi} \int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda \geq \sigma_0^2\,,$$

where σ_0^2 is the mean squared prediction error of the minimum mean squared error predictor of the TDGP, (Brockwell and Davis, 1991, Proposition 10.8.1). The implication of Assumption $A.7$ is that among all spectral densities within the mis-specified family the member characterized by the parameter value $\boldsymbol{\eta}_1$ is closest to the true spectral density. Evidently it is $\boldsymbol{\eta}_1$ that the estimators should be trying to target as this will give fitted parameter values that yield the predictor from within the MisM class whose mean squared prediction error is closest to that of the optimal predictor. Having established that the four parametric estimators converge towards $\boldsymbol{\eta}_1$ under mis-specification, we can as a consequence now broaden the applicability of the asymptotic distributional results derived by Chen and Deo (2006) for the FML estimator. This we do in Section 4 by establishing that all four alternative parametric estimators converge in distribution. Prior to doing this, however, we indicate the precise form of the limiting objective function $Q(\boldsymbol{\eta})$, and the associated first-order conditions that define the (common) pseudo-true value $\boldsymbol{\eta}_1$ of the four estimation procedures, in the ARFIMA case.

3 Pseudo-True Parameters Under ARFIMA Mis-Specification

Under Assumptions $A.1-A.7$, and under the condition that $(d_0-d) < 0.5$, $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ can be determined as the solution of the first-order condition $\partial Q(\boldsymbol{\eta})/\partial\boldsymbol{\eta} = 0$, and Chen and Deo (2006) illustrate the relationship between $\partial \log Q(\boldsymbol{\eta})/\partial d$ and the deviation $d^* = d_0 - d_1$ for the simple special case in which the TDGP is an $ARFIMA(0, d_0, 1)$ and the MisM is an $ARFIMA(0, d, 0)$. They then cite (without providing detailed derivations) certain results that obtain when the MisM is an $ARFIMA(1, d, 0)$. Here we provide a significant generalization, by deriving expressions for both $Q(\boldsymbol{\eta})$ and the first-order conditions that define the pseudo-true parameters, under the full $ARFIMA(p_0, d_0, q_0)/ARFIMA(p, d, q)$ dichotomy for the true process and the estimated model. Representations of the associated expressions via polynomial and power series expansions suitable for the analytical investigation of $Q(\boldsymbol{\eta})$ are presented. It is normally not possible to solve the first order conditions $\partial Q(\boldsymbol{\eta})/\partial\boldsymbol{\eta} = 0$ exactly as they are both nonlinear and (in general) defined as infinite sums. Instead one would determine the estimate numerically, via a Newton iteration for example, with the series expansions replaced by finite sums. An evaluation of the magnitude of the approximation error produced by any power series truncation that might arise from such a numerical implementation is given. The results are then illustrated in the special case where $p_0 = q = 0$, in which case true MA short memory dynamics of an arbitrary order are mis-specified as AR dynamics of an arbitrary order. In this particular case, as will be seen, no truncation error arises in the computations.

To begin, denote the spectral density of the TDGP, a general $ARFIMA(p_0, d_0, q_0)$ process, by

$$\frac{\sigma_0^2}{2\pi} f_0(\lambda) = \frac{\sigma_0^2}{2\pi} \frac{\left|1 + \theta_{10} e^{i\lambda} + ... + \theta_{q_0 0} e^{iq_0\lambda}\right|^2}{\left|1 + \phi_{10} e^{i\lambda} + ... + \phi_{p_0 0} e^{ip_0\lambda}\right|^2} |2\sin(\lambda/2)|^{-2d_0},$$

and that of the MisM, an $ARFIMA(p, d, q)$ model, by

$$\frac{\sigma^2}{2\pi} f_1(\boldsymbol{\eta}, \lambda) = \frac{\sigma^2}{2\pi} \frac{\left|1 + \theta_1 e^{i\lambda} + ... + \theta_q e^{iq\lambda}\right|^2}{\left|1 + \phi_1 e^{i\lambda} + ... + \phi_p e^{ip\lambda}\right|^2} |2\sin(\lambda/2)|^{-2d}.$$

Substituting these expressions into the limiting objective function we obtain the representation

$$Q(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2\pi} \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda = \frac{\sigma_0^2}{2\pi} \int_0^\pi \frac{|A_\beta(e^{i\lambda})|^2}{|B_\beta(e^{i\lambda})|^2} |2\sin(\lambda/2)|^{-2(d_0-d)} d\lambda\,, \tag{14}$$

where

$$A_\beta(z) = \sum_{j=0}^{\underline{q}} a_j z^j = \theta_0(z)\phi(z) = (1 + \theta_{10}z + ... + \theta_{q_0 0}z^{q_0})\,(1 + \phi_1 z + ... + \phi_p z^p), \tag{15}$$

with $\underline{q} = q_0 + p$ and

$$B_\beta(z) = \sum_{j=0}^{\underline{p}} b_j z^j = \phi_0(z)\theta(z) = (1 + \phi_{10}z + ... + \phi_{p_0 0}z^{p_0})\,(1 + \theta_1 z + ... + \theta_q z^q)\,, \tag{16}$$

with $\underline{p} = p_0 + q$. The expression for $Q(\boldsymbol{\eta})$ in (14) takes the form of the variance of an ARFIMA process with MA operator $A_\beta(z)$, AR operator $B_\beta(z)$ and fractional index $d_0 - d$. It follows that $Q(\boldsymbol{\eta})$ could be evaluated using the procedures presented in Sowell (1992). Sowell's algorithms are based upon series expansions in gamma and hypergeometric functions however, and although they are suitable for numerical calculations, they do not readily lend themselves to the analytical investigation of $Q(\boldsymbol{\psi})$. We therefore seek an alternative formulation.

Let $C(z) = \sum_{j=0}^{\infty} c_j z^j = A_\beta(z)/B_\beta(z)$ where $A_\beta(z)$ and $B_\beta(z)$ are as defined in (15) and (16) respectively. Then (14) can be expanded to give

$$Q\,(\boldsymbol{\eta}) = 2^{1-2(d_0-d)}\frac{\sigma_0^2}{2\pi}\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} c_j c_k \int_0^{\pi/2} \cos\left(2\,(j-k)\,\lambda\right)\sin(\lambda)^{-2(d_0-d)}d\lambda\right].$$

Using standard results for the integral $\int\limits_0^{\pi}(\sin x)^{\upsilon-1}\cos(ax)dx$ from Gradshteyn and Ryzhik (2007, p 397) yields, after some algebraic manipulation,

$$Q\,(\boldsymbol{\eta}) = \frac{\sigma_0^2}{2(1-2(d_0-d))}\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{c_j c_k \cos\left((j-k)\,\pi\right)}{\mathcal{B}\left(1-(d_0-d)+(j-k)\,,1-(d_0-d)-(j-k)\right)}\right],$$

where $\mathcal{B}(a,b)$ denotes the Beta function. This expression can in turn be simplified to

$$Q\,(\boldsymbol{\eta}) = \frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\}K(\boldsymbol{\eta})\,, \tag{17}$$

where

$$K(\boldsymbol{\eta}) = \sum_{j=0}^{\infty} c_j^2 + 2\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} c_j c_k \rho(j-k)$$

and

$$\rho(h) = \prod_{i=1}^{h}\left(\frac{(d_0-d)+i-1}{i-(d_0-d)}\right), \quad h = 1, 2, \ldots.$$

Using (17) we now derive the form of the first-order conditions that define $\boldsymbol{\eta}_1$, namely $\partial Q(\boldsymbol{\eta})/\partial\boldsymbol{\eta} = 0$. Differentiating $Q\,(\boldsymbol{\eta})$ first with respect to β_r, $r = 1, \ldots, l$, and then d gives:

$$\frac{\partial Q\,(\boldsymbol{\eta})}{\partial\beta_r} = \{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\}\frac{\partial K\,(\boldsymbol{\eta})}{\partial\beta_r}\,, \quad r = 1, 2, ..., l,$$

where

$$\frac{\partial K\,(\boldsymbol{\eta})}{\partial\beta_r} = \sum_{j=1}^{\infty} 2c_j\frac{\partial c_j}{\partial\beta_r} + 2\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty}(c_k\frac{\partial c_j}{\partial\beta_r} + \frac{\partial c_k}{\partial\beta_r}c_j)\rho(j-k)\,,$$

and

$$\frac{\partial Q(\boldsymbol{\eta})}{\partial d} = \{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\}\left\{2\left(\Psi[1-2(d_0-d)]-\Psi[1-(d_0-d)]\right)K(\boldsymbol{\eta})+\frac{\partial K(\boldsymbol{\eta})}{\partial d}\right\},$$

where $\Psi(\cdot)$ denotes the digamma function and

$$\begin{aligned}\frac{\partial K(\boldsymbol{\eta})}{\partial d} =& 2\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} c_j c_k \rho(j-k)\{2\Psi[1-(d_0-d)] \\ & -\Psi[1-(d_0-d)+(j-k)]-\Psi[1-(d_0-d)-(j-k)]\}\,.\end{aligned}$$

Eliminating the common (non-zero) factor $\{\pi\frac{\sigma_{\varepsilon 0}^2}{\sigma^2}\frac{\Gamma(1-2(d_0-d))}{\Gamma^2(1-(d_0-d))}\}$ from both $\partial Q(\boldsymbol{\eta})/\partial\boldsymbol{\beta}$ and $\partial Q(\boldsymbol{\eta})/\partial d$, it follows that the pseudo-true parameter values of the $ARFIMA(p,d,q)$ MisM can be obtained by solving

$$\frac{\partial K(\boldsymbol{\eta})}{\partial \beta_r} = 0\,, \quad r=1,2,...,l, \tag{18}$$

and

$$2(\Psi[1-2(d_0-d)]-\Psi[1-(d_0-d)])K(\boldsymbol{\eta})+\frac{\partial K(\boldsymbol{\eta})}{\partial d}=0 \tag{19}$$

for β_{r1}, $r=1,\ldots,l$, and d_1 using appropriate algebraic and numerical procedures. A corollary of the following theorem is that $\boldsymbol{\eta}_1$ can be calculated to any desired degree of numerical accuracy by truncating the series expansions in the expressions for $K(\boldsymbol{\eta})$, $\partial K(\boldsymbol{\eta})/\partial\boldsymbol{\beta}$ and $\partial K(\boldsymbol{\eta})/\partial d$ after a suitable number of N terms before substituting into (18) and (19) and solving (numerically) for ϕ_{i1}, $i=1,2,...,p$, θ_{j1}, $j=1,2,...,q$, and d_1.

Theorem 2 *Set $C_N(z)=\sum_{j=0}^{N}c_jz^j$ and let $Q_N(\boldsymbol{\eta})=\left(\sigma_0^2/\sigma_\varepsilon^2\right)I_N$ where the integral $I_N=\int_0^\pi|C_N(\exp(-i\lambda))|^2|2\sin(\lambda/2)|^{-2(d_0-d)}d\lambda$. Then*

$$Q(\boldsymbol{\eta})=Q_N(\boldsymbol{\eta})+R_N=\left\{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\right\}K_N(\boldsymbol{\eta})+R_N$$

where

$$K_N(\boldsymbol{\eta})=\sum_{j=0}^{N}c_j^2+2\sum_{k=0}^{N-1}\sum_{j=k+1}^{N}c_jc_k\rho(j-k)$$

and there exists a ζ, $0<\zeta<1$, such that $R_N=O(\zeta^{(N+1)})=o(N^{-1})$. Furthermore, $\partial Q_N(\boldsymbol{\eta})/\partial\boldsymbol{\eta}=\partial Q(\boldsymbol{\eta})/\partial\boldsymbol{\eta}+o(N^{-1})$.

By way of illustration, consider the case of mis-specifying a true $ARFIMA(0,d_0,q_0)$ process by an $ARFIMA(p,d,0)$ model. When $p_0=q=0$ we have $B_\beta(z)\equiv 1$ and $C(z)$ is polynomial, $C(z)=1+\sum_{j=1}^{\underline{q}}c_jz^j$ where $c_j=\sum_{r=\max\{0,j-p\}}^{\min\{j,p\}}\theta_{(j-r)0}\phi_r$. Abbreviating the latter to $\sum_r\theta_{(j-r)0}\phi_r$, this then gives us:

$$\begin{aligned}K(d,\phi_1,\ldots,\phi_p)=&\sum_{j=0}^{\underline{q}}(\sum_r\theta_{(j-r)0}\phi_r)^2+ \\ & 2\sum_{k=0}^{\underline{q}-1}\sum_{j=k+1}^{\underline{q}}(\sum_r\theta_{(j-r)0}\phi_r)(\sum_r\theta_{(k-r)0}\phi_r)\rho(j-k)\,;\end{aligned}$$

and setting $\theta_{s0}\equiv 0$, $s\ni[0,1,\ldots,q_0]$,

$$\begin{aligned}\frac{\partial K(d,\phi_1,\ldots,\phi_p)}{\partial\phi_r}=&\sum_{j=1}^{\underline{q}}2(\sum_r\theta_{(j-r)0}\phi_r)\theta_{(j-r)0}+ \\ & 2\sum_{k=0}^{\underline{q}-1}\sum_{j=k+1}^{\underline{q}}\left\{(\sum_r\theta_{(j-r)0}\phi_r)\theta_{(k-r)0}+\theta_{(j-r)0}(\sum_r\theta_{(k-r)0}\phi_r)\right\}\rho(j-k)\,,\end{aligned}$$

$r = 1, \ldots, p$, and

$$\frac{\partial K\left(d, \phi_1, \ldots, \phi_p\right)}{\partial d} = 2\sum_{k=0}^{\underline{q}-1}\sum_{j=k+1}^{\underline{q}}(\sum_r \theta_{(j-r)0}\phi_r)(\sum_r \theta_{(k-r)0}\phi_r)\rho(j-k) \times (2\Psi[1-(d_0-d)] - \Psi[1-(d_0-d)+(j-k)] - \Psi[1-(d_0-d)-(j-k)])$$

for the required derivatives. The pseudo-true values ϕ_{r1}, $r = 1, \ldots, p$, and d_1 can now be obtained by solving (18) and (19) having inserted these exact expressions for $K\left(d, \phi_1, \ldots, \phi_p\right)$, $\partial K\left(d, \phi_1, \ldots, \phi_p\right)/\partial\phi_r$, $r = 1, \ldots, p$, and $\partial K\left(d, \phi_1, \ldots, \phi_p\right)/\partial d$ into the equations.

Let us further highlight some features of this special case by focussing on the example where the TDGP is an $ARFIMA(0, d_0, 1)$ and the MisM an $ARFIMA(1, d, 0)$. In this example $\underline{q} = 2$ and $C(z) = 1 + c_1 z + c_2 z^2$ where, neglecting the first order MA and AR coefficient subscripts, $c_1 = (\theta_0 + \phi)$ and $c_2 = \theta_0\phi$. The second factor of the criterion function in (17) is now

$$K(d, \phi) = 1 + (\theta_0 + \phi)^2 + (\theta_0\phi)^2 + \frac{2\left[\theta_0\phi(d_0 - d + 1) - (1 + \theta_0\phi)(\theta_0 + \phi)(d_0 - d - 2)\right](d_0 - d)}{(d_0 - d - 1)(d_0 - d - 2)}. \tag{20}$$

The derivatives $\partial K(d, \phi)/\partial\phi$ and $\partial K(d, \phi)/\partial d$ can be readily determined from (20) and hence the pseudo-true values d_1 and ϕ_1 evaluated.

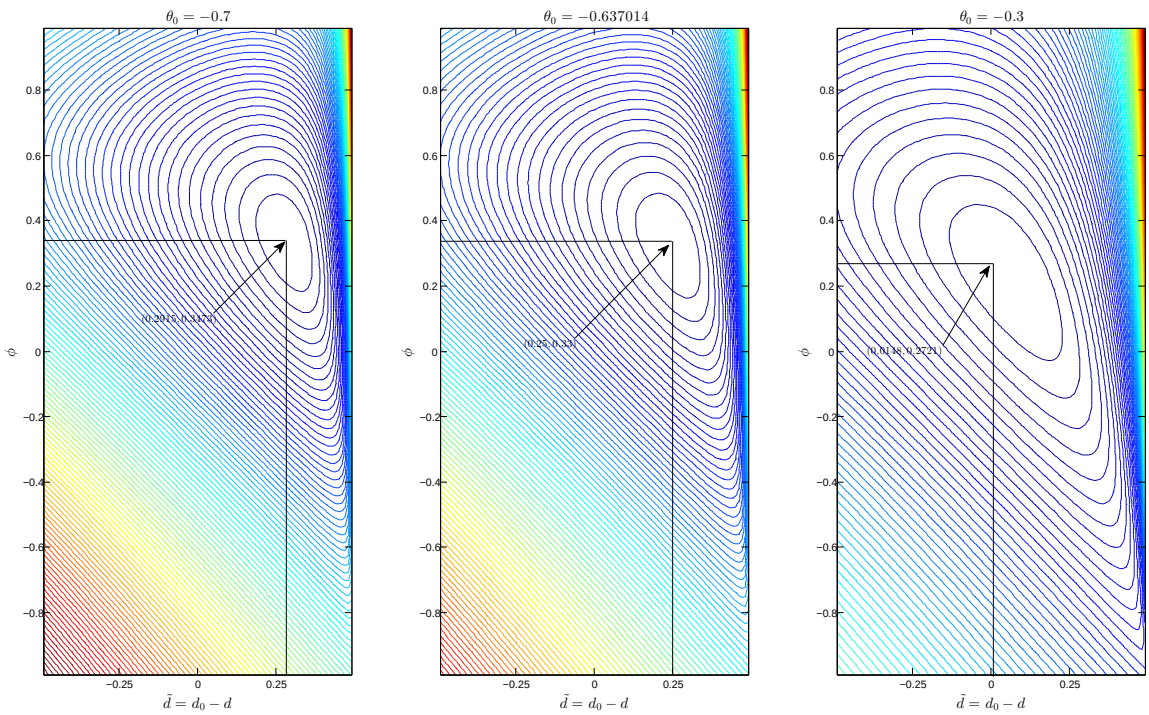
It is clear from (20) that for given values of $|\theta_0| < 1$ we can treat $K(d, \phi)$ as a function of $\widetilde{d} = (d_0 - d)$ and ϕ, and hence treat $Q\left(d, \phi\right) = Q\left(\boldsymbol{\eta}\right)$ similarly. Figure 1 depicts the contours of $Q\left(d, \phi\right)$ graphed as a function of $\widetilde{d}$ and ϕ for the values of $\theta_0 = \{-0.7, -0.637014, -0.3\}$ when $\sigma_\varepsilon^2 = \sigma_0^2$. Pre-empting the discussion to come in the following section, the values of θ_0 are deliberately chosen to coincide with $d^* = d_0 - d_1$ being respectively greater than, equal to and less than 0.25. The three graphs in Figure 1 clearly demonstrate the divergence in the asymptotic criterion function that occurs as $\widetilde{d} = (d_0 - d)$ approaches 0.5 and they illustrate that although the location of (d_1, ϕ_1) may be unambiguous, the sensitivity of $Q\left(d, \phi\right)$ to perturbations in (d, ϕ) can be very different depending on the value of $d^* = d_0 - d_1$.[2] In Figure 1a the contours indicate that when $d^* > 0.25$ the limiting criterion function has hyperbolic profiles in a small neighbourhood of the pseudo-true parameter point (d_1, ϕ_1), with similar but more locally quadratic behaviour exhibited in Figure 1b when $d^* = 0.25$. The contours of $Q(d, \phi)$ in Figure 1c, corresponding to $d^* < 0.25$, are more elliptical and suggest that in this case the limiting criterion function is far closer to being globally quadratic around (d_1, ϕ_1). It turns out that these three different forms of $Q\left(d, \phi\right)$, reflecting the most, intermediate, and the least mis-specified cases, correspond to the three different forms of asymptotic distribution presented in the following section.

4 Asymptotic Distributions

In this section we show that the asymptotic distribution of the FML estimator derived in Chen and Deo (2006) in the context of long range dependence is also applicable to the Whittle, TML and CSS estimators, and that all four estimators are, hence, asymptotically equivalent under mis-specification. As was highlighted by Chen and Deo, the rate of convergence and the nature of the asymptotic distribution of the FML estimator is determined by the deviation of the pseudo-true value of d, d_1, from the true value, d_0, with the three critical ranges for $d^* = d_0 - d_1 < 0.5$ being: $d^* > 0.25$, $d^* = 0.25$ and $d^* < 0.25$.[3] Theorem 3

[2] All the numerical results presented in this paper have been produced using MATLAB *2011b*, version *7.13.0.564 (R2011b)*.

[3] The results in Chen and Deo assumed that the parameter space of the estimated model coincided with the long memory region assumed for the TDGP. Also, since d_1 is only defined for $(d_0 - d) < 0.5$ it follows that the distributional results for the FML estimator were only valid for $d^* < 0.5$, something that was not explicitly mentioned by Chen and Deo in their original derivation.



(a) $\theta_0 = -0.7$. (b) $\theta_0 = -0.637014$. (c) $\theta_0 = -0.3$.

Figure 1: Contour plot of $Q(d, \phi)$ against $\widetilde{d} = d_0 - d$ and ϕ for the mis-specification of an ARFIMA$(0, d_0, 1)$ TDGP by an ARFIMA$(1, d, 0)$ MM; $\widetilde{d} \in (-0.5, 0.5)$, $\phi \in (-1, 1)$. Pseudo-true coordinates $(d_0 - d_1, \phi_1)$ are (a) $(0.2915, 0.3473)$, (b) $(0.25, 0.33)$ and (c) $(0.0148, 0.2721)$.

shows that in the event that any one of the FML, Whittle, TML or CSS estimators possesses the asymptotic distributions as described in the theorem, then all four estimators will share the same asymptotic distributions, and this will hold for all three forms of memory in the TDGP and the mis-specified model. We comment further on this matter below.

Theorem 3 *Suppose that the TDGP of $\{y_t\}$ is a Gaussian process with a spectral density as given in* (1) *and that the MisM in* (2) *satisfies Assumptions A.1 − A.7. Let*

$$\mathbf{B} = -\frac{\sigma_0^2}{\pi} \int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)} \frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}} \frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}^T} d\lambda + \frac{\sigma_0^2}{2\pi} \int_{-\pi}^{\pi} \frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)} \frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} d\lambda \,, \tag{21}$$

and set $\boldsymbol{\mu}_n = \mathbf{B}^{-1} E_0 \left(\frac{\partial Q_n(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} \right)$ where $Q_n(\cdot)$ denotes the objective function that defines $\widehat{\boldsymbol{\eta}}_1$.[4] *Assume that $(d_0 - d) < 0.5$ for all $\boldsymbol{\eta} \in \mathbb{E}$ and that $\boldsymbol{\eta}_1 \ni \partial\mathbb{E}$ where $\partial\mathbb{E}$ is the boundary of the set $\mathbb{E}$. Then if the limiting distribution of any one of the FML, Whittle, TML or CSS estimators is as delineated in Cases 1, 2 and 3, then the same is true of the other three estimators:*

Case 1: When $d^* = d_0 - d_1 > 0.25$,

$$\frac{n^{1-2d^*}}{\log n} \left(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1 - \boldsymbol{\mu}_n \right) \to^D \mathbf{B}^{-1} \left(\sum_{j=1}^{\infty} W_j, 0, ...0 \right)^T , \tag{22}$$

[4]Heuristically, $\boldsymbol{\mu}_n$ measures the bias associated with the estimator $\widehat{\boldsymbol{\eta}}_1$. That is, $\boldsymbol{\mu}_n \approx E_0 \left(\widehat{\boldsymbol{\eta}}_1 \right) - \boldsymbol{\eta}_1$. Note that the expression for $\boldsymbol{\mu}_n$ given in Chen and Deo (2006, p 263) contains a typographical error; the proofs in that paper use the correct expression. The derivation of $\boldsymbol{\mu}_n$ for all four estimation methods considered in this paper is provided in Appendix B.

where $\sum_{j=1}^{\infty} W_j$ *is defined as the mean-square limit of the random sequence* $\sum_{j=1}^{s} W_j$ *as* $s \to \infty$, *wherein*

$$W_j = \frac{(2\pi)^{1-2d^*}\, g_0(0)}{j^{2d^*} g_1(\boldsymbol{\beta},0)} \left[U_j^2 + V_j^2 - E_0(U_j^2 + V_j^2)\right],$$

and $\{U_j\}$ *and* $\{V_k\}$ *denote sequences of Gaussian random variables with zero mean and covariances* $Cov_0\,(U_j, U_k) = Cov_0\,(U_j, V_k) = Cov_0\,(V_j, V_k)$ *with*

$$Cov_0\,(U_j, V_k) = \iint_{[0,1]^2} \{\sin(2\pi jx)\sin(2\pi ky) + \sin(2\pi kx)\sin(2\pi jy)\}\, |x-y|^{2d_0-1}\, dxdy\,.$$

Case 2: When $d^* = d_0 - d_1 = 0.25$,

$$n^{1/2}\left[\overline{\Lambda}_{dd}\right]^{-1/2}(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1) \to^D \mathbf{B}^{-1}\,(Z, 0, ..., 0)^T\,, \tag{23}$$

where

$$\overline{\Lambda}_{dd} = \frac{1}{n}\sum_{j=1}^{n/2}\left(\frac{\sigma_0^2 f_0(\lambda_j)}{2\pi f_1(\boldsymbol{\eta}_1, \lambda_j)}\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda_j)}{\partial d}\right)^2,$$

and Z *is a standard normal random variable.*

Case 3: When $d^* = d_0 - d_1 < 0.25$,

$$\sqrt{n}\,(\widehat{\boldsymbol{\eta}}_1 - \boldsymbol{\eta}_1) \to^D N(0, \boldsymbol{\Xi}), \tag{24}$$

where $\boldsymbol{\Xi} = \mathbf{B}^{-1}\boldsymbol{\Lambda}\mathbf{B}^{-1}$,

$$\boldsymbol{\Lambda} = \frac{\sigma_0^4}{2\pi}\int_0^{\pi}\left(\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}_1,\lambda)}\right)^2\left(\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}}\right)\left(\frac{\partial \log f_1(\boldsymbol{\eta}_1,\lambda)}{\partial \boldsymbol{\eta}}\right)^T d\lambda\,.$$

A key point to note from the three cases delineated in Theorem 3 is that when the deviation between the true and pseudo-true values of d is sufficiently large ($d^* \geq 0.25$) – something that is related directly to the degree of mis-specification of $g_0(\lambda)$ by $g_1(\boldsymbol{\beta}, \lambda)$ – the $\sqrt{n}$ rate of convergence is lost, with the rate being arbitrarily close to zero depending on the value of d^*. For d^* strictly greater than 0.25, asymptotic Gaussianity is also lost, with the limiting distribution being a function of an infinite sum of non-Gaussian variables. For the $d^* \geq 0.25$ case, the limiting distribution – whether Gaussian or otherwise – is degenerate in the sense that the limiting distribution for each element of $\widehat{\boldsymbol{\eta}}_1$ is a different multiple of the same random variable ($\sum_{j=1}^{\infty} W_j$ in the case of $d^* > 0.25$ and Z in the case of $d^* = 0.25$).

For the form of limiting distribution that obtains in Cases 1, 2 and 3 we refer to Chen and Deo (2006, Theorems 1, 3 and 2), wherein these distributions were produced specifically for the FML estimator in the context of long range dependence. To prove that these same limiting distributions hold for the Whittle, TML and CSS estimators we establish that $R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(1)}) \to^D 0$ for $i = 2, 3$ and 4, where R_n denotes the convergence rate applicable in the three different cases outlined in the theorem. We use a first-order Taylor expansion of $\partial Q_n^{(\cdot)}(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}$ about $\partial Q_n^{(\cdot)}(\widehat{\boldsymbol{\eta}}_1^{(\cdot)})/\partial\boldsymbol{\eta} = \mathbf{0}$. This gives

$$\frac{\partial Q_n^{(\cdot)}(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}} = \frac{\partial^2 Q_n^{(\cdot)}(\grave{\boldsymbol{\eta}}_1^{(\cdot)})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\left(\boldsymbol{\eta}_1 - \widehat{\boldsymbol{\eta}}_1^{(\cdot)}\right)$$

and

$$R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(j)}) = \left[\frac{\partial^2 Q_n^{(j)}(\grave{\boldsymbol{\eta}}_1^{(j)})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right]^{-1} R_n\frac{\partial Q_n^{(j)}(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}} - \left[\frac{\partial^2 Q_n^{(i)}(\grave{\boldsymbol{\eta}}_1^{(i)})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right]^{-1} R_n\frac{\partial Q_n^{(i)}(\boldsymbol{\eta}_1)}{\partial\boldsymbol{\eta}}\,,$$

where $\|\boldsymbol{\eta}_1 - \grave{\boldsymbol{\eta}}_1^{(\cdot)}\| \leq \|\boldsymbol{\eta}_1 - \widehat{\boldsymbol{\eta}}_1^{(\cdot)}\|$. Since $\operatorname{plim} \widehat{\boldsymbol{\eta}}_1^{(\cdot)} = \boldsymbol{\eta}_1$ it is therefore sufficient to show that there exists a scalar, possibly constant, function $\mathcal{C}_n(\boldsymbol{\eta})$ such that

$$\frac{\partial^2 \{\mathcal{C}_n(\boldsymbol{\eta}_1) \cdot Q_n^{(i)}(\boldsymbol{\eta}_1) - Q_n^{(j)}(\boldsymbol{\eta}_1)\}}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} = o_p(1) \tag{25}$$

and

$$\operatorname{plim}_{n\to\infty} \quad R_n \left\| \mathcal{C}_n(\boldsymbol{\eta}_1) \cdot \frac{\partial Q_n^{(i)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} - \frac{\partial Q_n^{(j)}(\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}} \right\| = 0\,. \tag{26}$$

The condition in (25) is established by showing that for each $i = 1, 2, 3$ and 4 the Hessian $\partial^2\{Q_n^{(i)}(\boldsymbol{\eta}_1)\}/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ converges in probability to a matrix proportional to $\mathbf{B}$, as defined in (21). This result parallels the convergence of $Q_n^{(1)}(\boldsymbol{\eta})$ itself to the limiting objective function $Q(\boldsymbol{\eta})$ established in Proposition 1, following the replacement of $f_1(\boldsymbol{\eta}_1, \lambda)^{-1}$ by $\partial^2\{f_1(\boldsymbol{\eta}_1, \lambda)^{-1}\}/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ and $Q(\boldsymbol{\eta})$ by $\mathbf{B}$. The proof that the Hessians so converge uses arguments similar those employed in the proof of Proposition 1, the details are therefore omitted. The proof of (26) is more involved because of the presence of the scaling factor R_n. In Appendix A.5 we present the steps necessary to prove (26) for each estimator and TDGPs with fractional indices in the range $-0.5 < d_0 < 0.5$.

As noted above, Chen and Deo (2006) established that the three asymptotic distributions presented in Theorem 3 obtain for the FML estimator in the long memory case. Theorem 3 therefore implies that the same distributional properties hold true for the Whittle, TML and CSS estimators. It seems reasonable to conjecture that these asymptotic distributions will also hold true for the FML estimator in the short memory and antipersistent cases, and Theorem 3 would once again imply that the same distributional properties will hold true for the Whittle, TML and CSS estimators. Unfortunately a concise and neat proof that this conjecture is true currently evades us, although application of the arguments employed by Velasco and Robinson (2000) to extend the consistency and asymptotic normality of the FML estimator in correctly specified models of true fractional processes with indices that encompass the range $-0.5 < d_0 \leq 0$ seems likely to yield a suitable proof, and the simulation evidence cited in Section 5.2 serves as an indication of the conjecture's validity.

Finally, we highlight the fact that the FML and Whittle estimators are mean invariant by virtue of being defined on the non-zero fundamental Fourier frequencies. As a consequence, all convergence results presented in both this and the previous section for the FML and Whittle estimators also hold for a process that has an arbitrary (non-zero) mean, which may be unknown, thereby broadening the applicability of the theoretical results as they pertain to these particular estimators. The same is not true, however, either for the two time domain based methods or for the exact (integral-based) Whittle estimator, as will be demonstrated in Section 5.4 below.

5 Finite Sample Performance of the Mis-Specified Parametric Estimators of the Pseudo-True Parameter

5.1 Experimental design

In this section we explore the finite sample performance of the alternative methods, as it pertains to estimation of the pseudo-true value of the long memory parameter, d_1, under specific types of mis-specification. We give particular focus to the four estimators: $\widehat{d}_1^{(1)}$ (FML), $\widehat{d}_1^{(2)}$ (Whittle), $\widehat{d}_1^{(3)}$ (TML) and $\widehat{d}_1^{(4)}$ (CSS), for which the preceeding theoretical results have been produced. We first document the form of the finite sample distributions for each estimator by plotting the distribution of the standardized versions of the estimators, for which the asymptotic distributions are given in Cases 1, 2 and 3 respectively in Theorem 3. As part of this exercise we develop a method for obtaining the limiting distribution

for $d^* > 0.25$, as the distribution does not have a closed form in this case, as well as a method for estimating the bias-adjustment term, $\boldsymbol{\mu}_n$, which is relevant for this distribution. In the figures that follow the 'Limit' curve depicts the limiting distribution of the relevant statistic. Supplementing these graphical results, we then tabulate the bias and MSE of the four different techniques, as estimators of the pseudo-true parameter d_1, again under specific types of mis-specification and, hence, for different values of d^*. These results are supplemented by bias and MSE values for the exact Whittle estimator, which we refer to as $\widehat{d}_1^{(5)}$.

Data are simulated from a zero-mean Gaussian $ARFIMA(p_0, d_0, q_0)$ process, with the method of Sowell (1992), as modified by Doornik and Ooms (2001), used to compute the exact autocovariance function for the TDGP for any given values of p_0, d_0 and q_0. We have produced results for $n = 100$, 200, 500 and 1000 and for two versions of mis-specification nested in the general case for which the analytical results are derived in Section 3. However, we report selected results (only) from the full set due to space constraints. The bias and MSE results, plus certain computations needed for the numerical specification for the limiting distribution in the $d^* > 0.25$ case, are produced from $R = 1000$ replications of samples of size n from the relevant TDGP. The two forms of mis-specification considered are:

Example 1 *: An $ARFIMA(0, d_0, 1)$ TDGP, with parameter values $d_0 = \{-0.2, 0.2, 0.4\}$ and $\theta_0 = \{-0.7, -0.444978, -0.3\}$; and an $ARFIMA(0, d, 0)$ MisM. The value $\theta_0 = -0.7$ corresponds to the case where $d^* > 0.25$ and $\widehat{d}_1^{(i)}$, $i = 1, 2, 3, 4$, have the slowest rate of convergence, $n^{1-2d^*}/\log n$, and to a non-Gaussian distribution. The value $\theta_0 = -0.444978$ corresponds to the case where $d^* = 0.25$, in which case asymptotic Gaussianity is preserved but the rate of convergence is of order $(n/\log^3 n)^{1/2}$. The value $\theta_0 = -0.3$ corresponds to the case where $d^* < 0.25$, with $\sqrt{n}$-convergence to Gaussianity obtaining.*

Example 2 *: An $ARFIMA(0, d_0, 1)$ TDGP, with parameter values $d_0 = \{-0.2, 0.2, 0.4\}$ and $\theta_0 = \{-0.7, -0.637014, -0.3\}$; and an $ARFIMA(1, d, 0)$ MisM. In this example the value $\theta_0 = -0.7$ corresponds to the case where $d^* > 0.25$, the value $\theta_0 = -0.637014$ corresponds to the case where $d^* = 0.25$, and the value $\theta_0 = -0.3$ corresponds to the case where $d^* < 0.25$.*

In Subsection 5.2 we document graphically the form of the finite sampling distributions of all four estimators of d under the first type of mis-specification described above, and for $d_0 = 0.2$ only. The corresponding graphs under the different values of d_0 (and for all three cases) are qualitatively equivalent to those reported here and, hence, are not included.[5] However, and picking up on a point made above, the fact that the distributional results do not qualitatively alter when one allows for antiperistence in the TDPG (i.e. $d_0 = -0.2$) suggests that the three asymptotic distributions delineated in Theorem 3 continue to obtain for all four estimators beyond the long memory scenario that underpins the Chen and Deo (2006) distributional theory.

In Subsection 5.3 we report the bias and MSE of all four estimators (in terms of estimating the pseudo-true value d_1) under both forms of mis-specification and for all three values of d_0. Inclusion of finite sample bias and MSE results for negative d_0 is of interest given the theoretical validity of the consistency results in the antipersistence range. To supplement these results, all of which are based on the assumption that the mean is known, in Section 5.4 we reproduce corresponding bias and MSE results for the two time domain estimators, plus the exact Whittle estimator, using data in which the unknown mean is estimated along with the other parameters.

5.2 Finite sample distributions

In this section we consider in turn the three cases listed under Theorem 3. For notational ease and clarity we use $\widehat{d}_1$ to denote the (generic) estimator obtained under mis-specification,

[5] These additional graphical results are available from the authors on request.

remembering that this estimator may be produced by any one of the four estimation methods: $\widehat{d}_1^{(1)}$ to $\widehat{d}_1^{(4)}$. Similarly, we use $Q_n(\cdot)$ to denote the criterion associated with a generic estimator. Only when contrasting the (finite sample) performances of the alternative estimators do we re-introduce the superscript notation.

5.2.1 Case 1: $d^* > 0.25$

The limiting distribution for $\widehat{d}_1$ in this case is

$$\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right) \to^D b^{-1}\sum_{j=1}^{\infty} W_j\,, \tag{27}$$

where $\mu_n = b^{-1}E_0\left(\frac{\partial Q_n(\boldsymbol{\eta}_1)}{\partial d}\right)$,

$$\begin{aligned} b &= -2\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^3(\boldsymbol{\eta}_1,\lambda)}\left(\frac{\partial f_1(\boldsymbol{\eta}_1,\lambda)}{\partial d}\right)^2 d\lambda + \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1^2(\boldsymbol{\eta}_1,\lambda)}\frac{\partial^2 f_1(\boldsymbol{\eta}_1,\lambda)}{\partial d^2}d\lambda \\ &= -2\int_0^{\pi}(1+\theta_0^2+2\theta_0\cos(\lambda))(2\sin(\lambda/2))^{-2d^*}(2\log(2\sin(\lambda/2)))^2 d\lambda\,, \end{aligned} \tag{28}$$

and $W_j = \frac{(2\pi)^{1-2d^*}(1+\theta_0^2)}{j^{2d^*}}\left[U_j^2 + V_j^2 - E_0(U_j^2+V_j^2)\right]$, with $\{U_j\}$ and $\{V_k\}$ as defined in Theorem 3. (With reference to Theorem 3, both $\mathbf{B}$ and $\boldsymbol{\mu}_n$ in (22) are here scalars since in Example 1 there is only one parameter to estimate under the MisM, namely d. Hence the obvious changes made to notation. All other notation is as defined in the theorem.)

Given that the distribution in (27) is non-standard and does not have a closed form representation, consideration must be given to its numerical evaluation. In finite samples the bias-adjustment term μ_n (which approaches zero as $n \to \infty$) also needs to be calculated. We tackle each of these issues in turn, beginning with the computation of μ_n.

(1) From Theorem 3 it is apparent that in general the formula for $\mathbf{B}$ is independent of the estimation method, but the calculation of $\boldsymbol{\mu}_n$ requires separate evaluation of $E_0(\partial Q_n(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta})$ for each estimator. In Appendix B we provide expressions for $E_0(\partial Q_n(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta})$ for each of the four estimation methods. These formulae are used to evaluate the scalar μ_n here. Each value is then used in the specification of the standardized estimator $\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)$ in the simulation experiments.

(2) Quantification of the distribution of $\sum_{j=1}^{\infty} W_j$ requires the approximation of the infinite sum of the W_j, plus the use of simulation to represent the (appropriately truncated) sum. We truncate the series $\sum_{j=1}^{\infty} W_j$ after s terms where the truncation point s is chosen such that $1 \leqslant s < \lfloor n/2 \rfloor$ with $s \to \infty$ as $n \to \infty$ (*cf.* Lemma 6 of Chen and Deo (2006)). The value of s is determined using the following criterion function. Let

$$S_n = \widehat{Var}_0\left[\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)\right] \tag{29}$$

denote the empirical finite sample variation observed across the R replications and for each m, $1 \leqslant m < \lfloor n/2 \rfloor$, let

$$T_m = S_n - b^{-2}\Omega_m,$$

where $\Omega_m = Var_0\left(\sum_{j=1}^{m} W_j\right)$. Now set

$$s = \arg\min_{1\leqslant m<\lfloor n/2\rfloor} T_m. \tag{30}$$

Given s, we generate random draws of $\sum_{j=1}^{s} W_j$ via the underlying Gaussian random variables from which the W_j are constructed, and produce an estimate of the limiting distribution using kernel methods.

To determine s we need to evaluate

$$Var_0\left(\sum_{j=1}^{m} W_j\right) = \sum_{j=1}^{m} Var_0\left(W_j\right) + 2\sum_{j=1}^{m}\sum_{\substack{k=1\\ j\neq k}}^{m} Cov_0\left(W_j, W_k\right) . \tag{31}$$

The variance of W_j in this case is

$$\begin{aligned} &Var_0\left\{\frac{(2\pi)^{1-2d^*}\left(1+\theta_0^2\right)}{j^{2d^*}}\left[U_j^2+V_j^2-E_0\left(U_j^2+V_j^2\right)\right]\right\} \\ &= \frac{(2\pi)^{2-4d^*}\left(1+\theta_0^2\right)^2}{j^{4d^*}}\left\{E_0\left(U_j^2+V_j^2\right)^2-\left[E_0\left(U_j^2+V_j^2\right)\right]^2\right\}. \end{aligned}$$

As $\{U_j\}$ and $\{V_k\}$ are normal random variables with a covariance structure as specified in Theorem 3, standard formulae for the moments of Gaussian random variables yield the result that

$$\begin{aligned} E_0\left(U_j^2+V_j^2\right)^2 &= E_0\left(U_j^4\right)+2E_0\left(U_j^2V_j^2\right)+E_0\left(V_j^4\right) \\ &= 3\left[Var_0\left(U_j\right)\right]^2+2\left[Var_0\left(U_j\right)Var_0\left(V_j\right)+2Cov_0(U_j,V_j)\right] \\ &\quad +3\left[Var_0\left(V_j\right)\right]^2 \\ &= 12\left[Var_0\left(U_j\right)\right]^2 \end{aligned}$$

and

$$\begin{aligned} \left[E_0\left(U_j^2+V_j^2\right)\right]^2 &= \left[E_0\left(U_j^2\right)+E_0\left(V_j^2\right)\right]^2 \\ &= \left[Var_0\left(U_j\right)+Var_0\left(V_j\right)\right]^2 \\ &= 4\left[Var_0\left(U_j\right)\right]^2 . \end{aligned}$$

Thus,

$$Var_0(W_j) = \frac{8\,(2\pi)^{2-4d^*}\left(1+\theta_0^2\right)^2}{j^{4d^*}}\left[Var_0\left(U_j\right)\right]^2 .$$

Similarly, the covariance between W_j and W_k when $j \neq k$ can be shown to be equal to

$$\begin{aligned} &\frac{(2\pi)^{2-4d^*}\left(1+\theta_0^2\right)^2}{(jk)^{2d^*}}Cov_0\left(U_j^2+V_j^2, U_k^2+V_k^2\right) \\ &\qquad = \frac{4\,(2\pi)^{2-4d^*}\left(1+\theta_0^2\right)^2}{(jk)^{2d^*}}\left[Var_0\left(U_j\right)Var_0\left(V_k\right)+2Cov_0(U_j,V_k)\right] . \end{aligned}$$

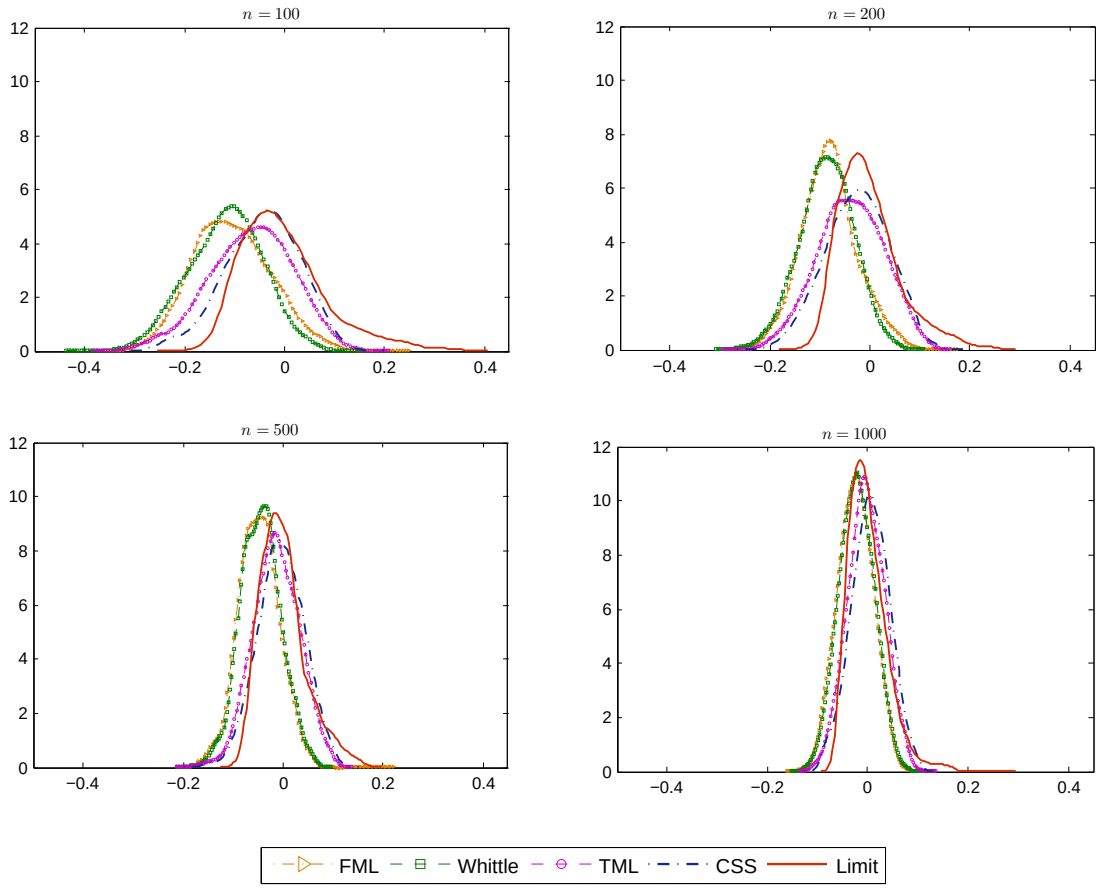
The expression in (31) can therefore be evaluated numerically using the formula for $Cov_0(U_j, V_k)$ to calculate the necessary moments required to determine s from (30).

The idea behind the use of T_m is simply to minimize the difference between the second-order sample and population moments. The value of S_n in (29) will vary with the estimation method of course; however, we choose s based on S_n calculated from the FML estimates and maintain this choice of s for all other methods. The terms in (31) are also dependent on the form of both the TDGP and the MisM and hence T_m needs to be determined for any specific case. The values of s for the sample sizes used in the particular simulation experiment underlying the results in this section are provided in Table 1.

Table 1 Truncation values s corresponding to the case $d^*= 0.3723$ for the Ex 1: $ARFIMA(0, d_0, 1)$ TDGP with $d_0= 0.2$ and $\theta_0= -0.7$ vis-à-vis MisM: $ARFIMA(0, d, 0)$.

n	100	200	500	1000
s	36	75	162	230

Figure 2: Kernel density of $\frac{n^{1-2d^*}}{\log n}\left(\widehat{d}_1 - d_1 - \mu_n\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0= 0.2$ and $\theta_0= -0.7$, and an $ARFIMA(0, d, 0)$ MM; $d^*> 0.25$.



Each panel in Figure 2 provides the kernel density estimate of $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ under the four estimation methods, for a specific n as labeled above each plot, plus the limiting distribution for the given s. The particular parameter values employed in the specification of the TGDP are $d_0 = 0.2$ and $\theta_0 = -0.7$, with $d^* = 0.3723$ in this case, and the values of s used are those given in Table 1. From Figure 2 we see that $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ is centered away from zero for all sample sizes, for all estimation methods. However, as the sample size increases the point of central location of $\frac{n^{1-2d^*}}{\log n}(\widehat{d}_1 - d_1 - \mu_n)$ approaches zero and all distributions of the standardized statistics go close to matching the asymptotic ('limit') distributions. The salient feature to be noted is the clustering that occurs, in particular for $n \leqslant 500$; that is, TML and CSS form one cluster and FML and Whittle form the other, with the time-domain estimators being closer to the asymptotic distribution for all three (smaller) sample sizes.

5.2.2 Case 2: $d^* = 0.25$

The limiting distribution for $\widehat{d}_1$ in the case of $d^* = 0.25$ is

$$n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right) \to^D N(0, b^{-2}), \tag{32}$$

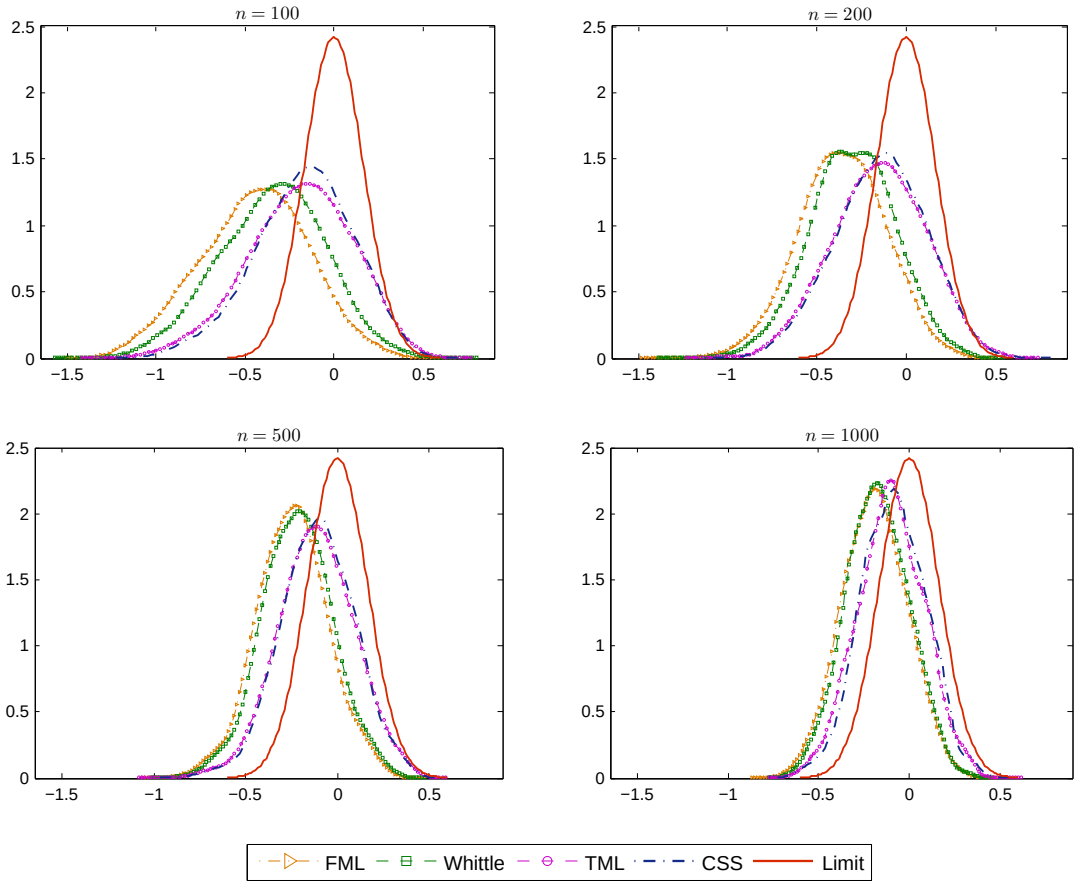
where

$$\overline{\Lambda}_{dd} = \frac{1}{n}\sum_{j=1}^{n/2}(1+\theta_0^2+2\theta_0\cos(\lambda_j))^2(2\sin(\lambda_j/2))^{-1}(2\log(2\sin(\lambda_j/2)))^2 \tag{33}$$

and b is as in (28). In both (33) and (28) $\theta_0 = -0.444978$, as $d^* = 0.25$ occurs at this particular value. Once again, $d_0 = 0.2$ in the TDGP.

Each panel of Figure 3 provides the densities of $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right)$ under the four estimation methods, for a specific n as labeled above each plot, plus the limiting distribution given in (32). Once again we observe a disparity between the time domain and frequency

Figure 3: Kernel density of $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with d_0= 0.2 and θ_0= −0.444978, and an $ARFIMA(0, d, 0)$ MM, d^*= 0.25.



domain kernel estimates, with the pair of time domain methods yielding finite sample distributions that are closer to the limiting distribution, for all sample sizes considered. The discrepancy between the two types of methods declines as the sample size increases, with the distributions of all methods being reasonably close both to one another, and to the limiting distribution, when $n = 1000$.

5.2.3 Case 3: $d^* < 0.25$

In this case we have

$$\sqrt{n}\left(\widehat{d}_1 - d_1\right) \to^D N(0, \upsilon^2)\,, \tag{34}$$

where

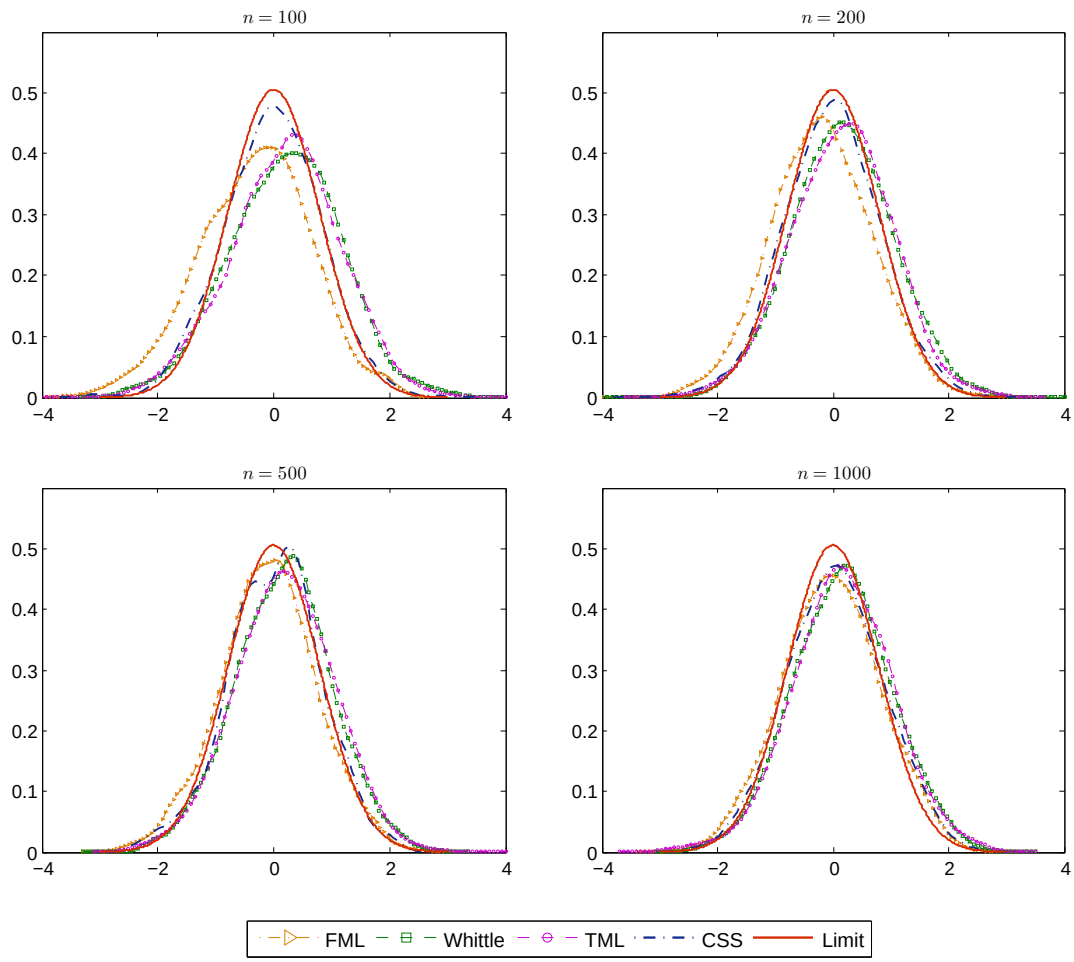
$$\upsilon^2 = \Lambda_{11}/b^{-2}\,, \tag{35}$$

with

$$\begin{aligned}\Lambda_{11} &= 2\pi\frac{\sigma_0^2}{2\pi}\int_0^\pi \left(\frac{f_0(\lambda)}{f_1(d_1,\lambda)}\right)^2 \left(\frac{\partial \log f_1(d_1,\lambda)}{\partial d}\right)^2 d\lambda \\ &= 2\pi\int_0^\pi (1+\theta_0^2+2\theta_0\cos(\lambda))^2(2\sin(\lambda/2))^{-4d^*}(2\log(2\sin(\lambda/2)))^2 d\lambda\,,\end{aligned}$$

and b as given in (28) evaluated at $\theta_0 = -0.3$ and $d^* = 0.1736$. Each panel in Figure 4 provides the kernel density estimate of the standardized statistic $\sqrt{n}(\widehat{d}_1 - d_1)$, under the four estimation methods, for a specific n as labeled above each plot, plus the limiting distribution given in (34). In this case there is no clear visual differentiation between the time domain

Figure 4: Kernel density of $\sqrt{n}\left(\widehat{d}_1 - d_1\right)$ for an $ARFIMA(0, d_0, 1)$ TDGP with $d_0 = 0.2$ and $\theta_0 = -0.3$, and an $ARFIMA(0, d, 0)$ MM, $d^* < 0.25$.



and frequency domain methods, for any sample size, and perhaps not surprisingly given the faster convergence rate in this case, all the methods produce finite sample distributions that match the limiting distribution reasonably well by the time $n = 1000$.

5.3 Finite sample bias and MSE of estimators of the pseudo-true parameter d_1 : known mean case

We supplement the graphical results in the previous section by documenting the finite sample bias and MSE of the four alternative estimators discussed in the previous section, in addition to the exact Whittle estimator, as estimators of the pseudo-true parameter d_1. The following standard formulae,

$$\widehat{\text{Bias}}_0\left(\widehat{d}_1^{(i)}\right) = \frac{1}{R}\sum_{r=1}^{R}\widehat{d}_r^{(i)} - d_1 \tag{36}$$

$$\widehat{Var}_0\left(\widehat{d}_1^{(i)}\right) = \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{d}_{1,r}^{(i)}\right)^2 - \left(\frac{1}{R}\sum_{r=1}^{R}\widehat{d}_{1,r}^{(i)}\right)^2 \tag{37}$$

$$\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(i)}\right) = \widehat{\text{Bias}}_0^2 + \widehat{Var}_0\left(\widehat{d}_1^{(i)}\right) \tag{38}$$

$$\widehat{r.eff}_0\left(\widehat{d}_1^{(i)}, \widehat{d}_1^{(j)}\right) = \frac{\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(i)}\right)}{\widehat{\text{MSE}}_0\left(\widehat{d}_1^{(j)}\right)}, \tag{39}$$

are applied to all five estimators $i, j = 1, ..., 5$. Since all empirical expectations and variances are evaluated under the TDGP, we make this explicit with appropriate subscript notation. Results for known mean are produced for Example 1 and Example 2 in Table 2 and 3 respectively, with selected additional results relevant to both examples recorded in Table 4. Values of $d^* = d_0 - d_1$ are documented across the key ranges, $d^* \lesseqgtr 0.25$, along with associated values for the MA coefficient in the TDGP, θ_0. The minimum values of bias and MSE for each parameter setting are highlighted in bold face in all tables for each sample size, n.[6] Corresponding results for the case in which the mean is estimated are recorded in Section 5.4.

Consider first the bias and MSE results for Example 1 with $d_0 = 0.2$, as displayed in the middle panel of Table 2. As is consistent with the theoretical results (and the graphical illustration in the previous section) the bias and MSE of the four parametric estimators FML, Whittle, TML and CSS, show a clear tendency to decline as the sample size increases, for a fixed value of θ_0. In addition, as θ_0 declines in magnitude, and the MisM becomes closer to the TDGP, there is a tendency for the MSE values and the absolute values of the bias to decline. Importantly, the bias is *negative* for all four estimators, with the (absolute) bias of the two frequency domain estimators (FML and Whittle) being larger than that of the two time domain estimators. These results are consistent with the tendency of the standardized sampling distributions illustrated above to cluster, and for the frequency domain estimators to sit further to the left of zero than those of the time domain estimators, at least for the $d^* \geq 0.25$ cases. Again, as is consistent with the theoretical results, the rate of decline in the (absolute) bias and MSE of all estimators, as n increases, is slower for $d^* \geq 0.25$ than for $d^* < 0.25$. The performance of the exact Whittle estimator (as we term it) is (almost) uniformly better than that of the Whittle estimator, but remains inferior to that of the two time domain estimators, with the CSS being clearly the superior estimator overall. The exact Whittle procedure mimics the other four methods in terms of the decline in both (absolute) bias and MSE as n increases, providing numerical evidence that this frequency domain estimator is also consistent for d_1.[7]

As indicated by the results in the bottom panel of Table 2 for $d_0 = 0.4$, the impact of an increase in d_0 (for any given value of d^* and n) is to often (but not uniformly) increase

[6] Only that number which is smallest at the precision of 8 decimal places is bolded. Values highlighted with a '$*$' are equally small to 4 decimal places.

[7] The tendency for the exact Whittle estimator to have (in particular) smaller finite sample bias than its inexact counterpart would seem to confirm the speculation in (Chen and Deo, 2006, Section 2) that the bias term ($\boldsymbol{\mu}_n$ in Theorem 3) may converge to zero more quickly for the first estimator than for the second.

Table 2: Estimates of the bias and MSE of $\widehat{d_1}$ for the FML, Whittle, Exact Whittle, TML and CSS estimators corresponding to Example 1 - TDGP: $ARFIMA(0, d_0, 1)$ vis-à-vis Mis-M: $ARFIMA(0, d, 0)$. Process mean, $\mu = 0$, is known.

				FML		Whittle		Exact Whittle		TML		CSS	
d^*	d_0	θ_0	n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.3723	-0.2	-0.7	100	-0.2416	0.0698	-0.2085	0.0578	-0.1317	0.0388	-0.2055	0.0542	**-0.0941**	**0.0216**
			500	-0.1332	0.0209	-0.1243	0.0185	-0.1107	0.0128	-0.1175	0.0168	**-0.0631**	**0.0081**
			1000	-0.1053	0.0129	-0.1018	0.0121	-0.1053	0.0115	-0.0984	0.0113	**-0.0465**	**0.0050**
0.2500		-0.44	100	-0.1425	0.0308	-0.1068	0.0227	-0.0823	0.0170	-0.0837	0.0169	**-0.0401**	**0.0109**
			500	-0.0539	0.0048	-0.0502	0.0045	-0.0460	0.0039	-0.0414	0.0037	**-0.0192**	**0.0026**
			1000	-0.0378	0.0024	-0.0371	0.0024	-0.0308	0.0022	-0.0325	0.0021	**-0.0092**	**0.0014**
0.1736		-0.3	100	-0.1071	0.0219	-0.0645	0.0147	-0.0468	0.0156	-0.0464	0.0111	**-0.0215**	**0.0087**
			500	-0.0345	0.0028	-0.0272	0.0025	-0.0154	0.0020	-0.0212	0.0021	**-0.0087**	**0.0018**
			1000	-0.0218	0.0013	-0.0191	0.0012	-0.0167	0.0011	-0.0156	0.0010	**-0.0023**	**0.0009**
0.3723	0.2	-0.7	100	-0.1781	0.0915	-0.2466	0.0691	-0.1968	0.0550	-0.1748	0.0481	**-0.1427**	**0.0315**
			500	-0.1354	0.0211	-0.1308	0.0178	-0.1017	0.0141	-0.0916	0.0138	**-0.0798**	**0.0097**
			1000	-0.1019	0.0141	-0.0996	0.0127	-0.0789	0.0097	-0.0776	0.0103	**-0.0670**	**0.0065**
0.2500		-0.44	100	-0.1515	0.0393	-0.1184	0.0298	-0.0795	0.0192	-0.0650	0.0170	**-0.0577**	**0.0119**
			500	-0.0544	0.0048	-0.0487	0.0042	-0.0307	0.0035	-0.0257	0.0027	**-0.0241**	**0.0021**
			1000	-0.0351	0.0023	-0.0323	0.0021	-0.0233	0.0018	-0.0188	0.0015	**-0.0162**	**0.0012**
0.1736		-0.3	100	-0.1082	0.0217	-0.0712	0.0146	-0.0428	0.0109	-0.0340	0.0095	**-0.0330**	**0.0087**
			500	-0.0318	0.0026	-0.0251	0.0022	-0.0184	0.0018	**-0.0106**	0.0017*	-0.0188	**0.0017**
			1000	-0.0184	0.0011	-0.0149	0.0010	-0.0077	0.0010	**-0.0065**	0.0009*	-0.0180	**0.0009**
0.3723	0.4	-0.7	100	-0.2786	0.0995	-0.2456	0.0724	-0.2165	0.0586	-0.2210	0.0515	**-0.1957**	**0.0489**
			500	-0.1598	0.0213	-0.1287	0.0181	-0.1142	0.0140	-0.1347	0.0137	**-0.0871**	**0.0118**
			1000	-0.1123	0.0157	-0.0939	0.0143	-0.0821	0.0113	-0.0812	0.0121	**-0.0648**	**0.0117**
0.2500		-0.44	100	-0.1903	0.0475	-0.1659	0.0383	-0.1086	0.0214	-0.0911	0.0201	**-0.0550**	**0.0138**
			500	-0.0796	0.0095	-0.0550	0.0082	-0.0303	0.0062	-0.0249	0.0059	**-0.0224**	**0.0038**
			1000	-0.0360	0.0048	-0.0295	0.0042	-0.0202	0.0038	-0.0180	0.0035	**-0.0175**	**0.0025**
0.1736		-0.3	100	-0.0990	0.0228	-0.0843	0.0152	-0.0509	0.0122	-0.0422	0.0102	**-0.0321**	**0.0092**
			500	-0.0407	0.0031	-0.0276	0.0025	-0.0188	0.0025	-0.0129	0.0022	**-0.0087**	**0.0019**
			1000	-0.0172	0.0011	-0.0163	0.0010	-0.0089	0.0009	-0.0077	0.0009	**-0.0052**	**0.0008**

the bias and MSE of all (five) estimator of d_1. That is, the ability of the estimators to accurately estimate the pseudo-true parameter tends to decline (overall) as the long memory in the TDGP increases. In contrast, and with reference to the results in the top panel of the table for the antipersistent case, estimation accuracy tends to increase (as a general rule) as the memory in the TDGP declines. Nevertheless, the results in Table 2 show that the *relativities* between the estimators remain essentially the same for the different values of d_0, with the CSS estimator remaining preferable overall to all other estimators under mis-specification, and the FML estimator performing the worst of all.[8]

The results recorded in Table 3 for Example 2 illustrate that the presence of an AR term in the MisM means that more severe mis-specification can be tolerated. More specifically, in all (comparable) cases and for all estimators, the finite sample bias and MSE recorded in Table 3 tend to be smaller in (absolute) value than the corresponding values in Table 2. Results not presented here suggest, however, that when the value of θ_0 is near zero, estimation under the MisM with an extraneous AR parameter causes an increase in (absolute) bias and MSE, relative to the case where the MisM is fractional noise (see also the following remark). With due consideration taken of the limited nature of the experimental design, these results suggest that the inclusion of some form of short memory dynamics in the estimated model – even if those dynamics are not of the correct form – acts as an insurance against more extreme mis-specification, but at the possible cost of a decline in performance when the consequences of mis-specification are not severe.

REMARK: When the parameter θ_0 of the $ARFIMA(0, d_0, 1)$ TDGP equals zero the TDGP coincides with the $ARFIMA(0, d, 0)$ model and is nested within the $ARFIMA(1, d, 0)$ model. Thus the value $\theta_0 = 0$ is associated with a match between the TDGP and the model, at which point $d^* = 0$ and there is no mis-specification. That is, neither the $ARFIMA(0, d, 0)$ model estimated in Example 1, nor the $ARFIMA(1, d, 0)$ model estimated in Example 2, is mis-specified (according to our definition) when applied to an ARFIMA $(0, d_0, 0)$ TDGP, although the $ARFIMA(1, d, 0)$ model is incorrect in the sense of being over-parameterized. Table 4 presents the bias and MSE observed when there is such a lack of mis-specification. Under the correct specification of the $ARFIMA(0, d, 0)$ model the TML estimator is now superior, in terms of both bias and MSE. The relative accuracy of the TML estimator seen here is consistent with certain results recorded in Sowell (1992) and Cheung and Diebold (1994), in which the performance of the TML method (under a known mean, as is the case considered here) is assessed against that of various comparators under correct model specification. For the over-parameterized $ARFIMA(1, d, 0)$ model, however, the CSS estimator dominates once more. □

The results in Tables 2, 3 and 4 highlight that, in all but one case, the CSS estimator has the smallest MSE of all five estimators under mis-specification, and when there is no mis-specification but the model is over-parameterized, and that this result holds for all sample sizes considered. Indeed, the MSE results indicate that the CSS estimator is between about two and three times as efficient as the FML estimator (in particular) in the region of the parameter space ($d^* \geq 0.25$) in which both (absolute) bias and MSE are at their highest for all estimators. The absolute value of the bias of CSS is also the smallest in the vast majority of such cases, for all values of d^*. This almost universal superiority of the CSS method presumably reflects a certain in-built robustness of least squares methods.

Table 3: Estimates of the bias and MSE of $\widehat{d_1}$ for the FML, Whittle, Exact Whittle, TML and CSS estimators corresponding to Example 2 - TDGP: $ARFIMA(0, d_0, 1)$ vis-à-vis Mis-M: $ARFIMA(1, d, 0)$. Process mean, $\mu = 0$, is known.

				FML		Whittle		Exact Whittle		TML		CSS	
d^*	d_0	θ_0	n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.2915	-0.2	-0.7	100	-0.1174	0.0503	-0.0809	0.0216	-0.0756	0.0197	-0.0632	0.0163	**-0.0526**	**0.0138**
			500	-0.0210	0.0115	-0.0179	0.0109	-0.0153	0.0098	-0.0134	0.0085	**-0.0122**	**0.0082**
			1000	-0.0105	0.0054	-0.0088	0.0049	-0.0079	0.0043	-0.0070	0.0044	**-0.0053**	**0.0042**
0.25		-0.64	100	-0.0868	0.0198	-0.0770	0.0153	-0.0641	0.0129	-0.0507	0.0106	**-0.0420**	**0.0096**
			500	-0.0274	0.0039	-0.0167	0.0026	-0.0135	0.0031	**-0.0113**	**0.0021**	-0.0121	0.0023
			1000	-0.0166	0.0012	-0.0128	0.0011	-0.0101	0.0010	-0.0094	0.0010	**-0.0085**	**0.0010**
0.0148		-0.3	100	-0.0320	0.0123	-0.0242	0.0099	-0.0192	0.0103	-0.0156	0.0069	**-0.0095**	**0.0065**
			500	-0.0161	0.0025	-0.0137	0.0031	-0.0089	0.0029	-0.0064	0.0013	**-0.0030**	**0.0013**
			1000	-0.0094	0.0009	-0.0075	0.0007	-0.0055	0.0007	-0.0037	0.0007*	**-0.0011**	**0.0007**
0.2915	0.2	-0.7	100	-0.1612	0.0541	-0.1169	0.0342	-0.0982	0.0304	-0.0950	0.0295	**-0.0671**	**0.0236**
			500	-0.0679	0.0165	-0.0604	0.0125	-0.0557	0.0113	-0.0454	0.0110	**-0.0369**	**0.0089**
			1000	-0.0469	0.0089	-0.0432	0.0071	-0.0339	0.0070	**-0.0250**	0.0067	-0.0303	**0.0059**
0.25		-0.64	100	-0.1339	0.0279	-0.0899	0.0175	-0.0764	0.0145	-0.0655	0.0138	**-0.0457**	**0.0110**
			500	-0.0490	0.0041	-0.0415	0.0030	-0.0373	0.0032	-0.0323	0.0026	**-0.0230**	**0.0022**
			1000	-0.0316	0.0019	-0.0281	0.0015	-0.0228	0.0016	-0.0181	0.0013	**-0.0176**	**0.0011**
0.0148		-0.3	100	-0.0508	0.0139	-0.0256	0.0086	-0.0206	0.0075	0.0190	0.0067	**-0.0082**	**0.0054**
			500	-0.0093	0.0027	-0.0080	0.0019	-0.0078	0.0018	0.0073	0.0016	**-0.0004**	**0.0014**
			1000	-0.0036	0.0010	-0.0023	0.0008	-0.0020	0.0007	0.0067	0.0006*	**0.0003**	**0.0006**
0.2915	0.4	-0.7	100	-0.2299	0.0639	-0.1805	0.0419	-0.1483	0.0389	-0.1279	0.0372	**-0.0699**	**0.0140**
			500	-0.1294	0.0197	-0.1039	0.0150	-0.0948	0.0134	-0.0816	0.0126	**-0.0294**	**0.0101**
			1000	-0.1089	0.0125	-0.0632	0.0099	-0.0521	0.0084	-0.0462	0.0081	**-0.0109**	**0.0069**
0.25		-0.64	100	-0.1396	0.0257	-0.0979	0.0155	-0.0804	0.0148	-0.0692	0.0145	**-0.0508**	**0.0103**
			500	-0.0455	0.0065	-0.0342	0.0046	-0.0315	0.0052	-0.0294	0.0041	**-0.0216**	**0.0033**
			1000	-0.0316	0.0027	-0.0192	0.0021	-0.0156	0.0020	-0.0177	0.0018	**-0.0122**	**0.0014**
0.0148		-0.3	100	-0.0650	0.0162	-0.0422	0.0115	-0.0368	0.0105	0.0246	0.0082	**-0.0132**	**0.0067**
			500	-0.0205	0.0042	-0.0133	0.0034	-0.0099	0.0031	0.0079	0.0026	**-0.0035**	**0.0023**
			1000	-0.0136	0.0021	-0.0088	0.0018	-0.0065	0.0015	0.0053	0.0014	**-0.0017**	**0.0013**

Table 4 Estimates of the bias and MSE of $\widehat{d}_1$ for the FML, Whittle, Exact Whittle, TML and CSS estimators corresponding to TDGP: $ARFIMA(0, d_0, 0)$ $d_0 = 0.2$, $d^* = 0.0$ with the known process mean, $\mu = 0$.

	FML		Whittle		Exact Whittle		TML		CSS	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
					Correct $ARFIMA(0, d, 0)$model					
100	-0.0502	0.0113	-0.0173	0.0102	-0.0089	0.0099	**0.0066**	**0.0087**	0.0094	0.0096
500	-0.0089	0.0015	-0.0062	0.0014	-0.0033	0.0014	**0.0026**	**0.0013**	0.0031	0.0014
1000	-0.0045	0.0006*	-0.0037	0.0006*	-0.0028	0.0009	**0.0016**	**0.0006**	0.0025	0.0006*
					Over-Parameterized $ARFIMA(1, d, 0)$model					
100	-0.0455	0.0177	0.0371	0.0121	-0.0364	0.0122	0.0255	0.0107	**0.0158**	**0.0087**
500	-0.0120	0.0065	0.0091	0.0049	0.0083	0.0039	0.0078	0.0043	**0.0055**	**0.0037**
1000	-0.0074	0.0027	0.0055	0.0021	0.0040	0.0020	0.0034	0.0019	**0.0028**	**0.0016**

5.4 Finite sample bias and MSE of estimators of the pseudo-true parameter d_1 : unknown mean case

Since the FML and Whittle estimators are both mean invariant the results recorded in the previous two sections for these two estimators are applicable to the unknown (zero) mean case without change. What will potentially alter, however, will be the performance of these two frequency domain estimators relative to that of the exact Whittle and time domain estimators when the unknown mean is also estimated, and it is that possibility that we explore in this section.

In Table 5 we record the bias and MSE obtained for the exact Whittle, TML and CSS estimators when the true mean of the process, μ, is estimated using the sample mean; for both mis-specified examples, all three values of d_0, and all three sample sizes. Properties observed in the previous section, such as the decline in bias and MSE with an increase in sample size, for a given θ_0, and the overall decline in MSE and (absolute) bias as the estimated model becomes less mis-specified, continue to obtain in Table 5. However, the magnitudes of the bias and MSE figures for all three estimators are now virtually always higher than the corresponding figures in Tables 2 and 3. As a consequence of this, the time domain estimators lose their relative superiority and no longer uniformly dominate the frequency domain techniques. Instead, the Whittle estimator outperforms all four of the other estimators overall (including its exact counterpart), and almost uniformly in the (true) long memory cases ($d_0 = 0.2, 0.4$). Table 6 records the outcomes obtained for the exact Whittle, TML and CSS estimators under the correct and over-parameterized specifications when the mean is estimated. Comparing Table 6 with Table 4 we find (once again) that the Whittle estimator now dominates all other estimators. As the sample size increases the differences between all comparable results for the known and estimated mean cases become less marked, in accordance with the consistency of the estimated mean for the true (zero) mean.[9]

[8] A slight caveat to this statement is that the superiority of the TML estimator over the exact Whittle is slightly less uniform when either $d = -0.2$ or $d = 0.4$. The overall similarity of the performance of these two estimators across the whole parameter space is, however, not surprising.

[9] The results recorded here regarding the performance of the different estimators in the unknown mean case parallel the qualitative conclusions drawn by Nielsen and Frederiksen (2005) for correctly specified models.

Table 5: Estimates of the bias and MSE of $\widehat{d}_1$ for the Exact Whittle, TML and CSS estimators when the mean is also estimated using the sample mean. The values of d^* for each design are given in Table 2.

		Example 1: TDGP: $ARFIMA(0, d_0, 1)$; MisM: $ARFIMA(0, d, 0)$							Example 2: TDGP: $ARFIMA(0, d_0, 1)$; MisM: $ARFIMA(1, d, 0)$						
			Exact Whittle		TML		CSS			Exact Whittle		TML		CSS	
n	d_0	θ_0	Bias	MSE	Bias	MSE	Bias	MSE	θ_0	Bias	MSE	Bias	MSE	Bias	MSE
100	-0.2	-0.7	-0.1824	0.0468	-0.2139	0.0581	-0.1543	0.0318	-0.7	-0.1078	0.0356	-0.0951	0.0364	-0.0704	0.0269
500			-0.1223	0.0254	-0.1286	0.0193	-0.0965	0.0122		-0.0487	0.0113	-0.0332	0.0101	-0.0265	0.0096
1000			-0.0823	0.0148	-0.1086	0.0133	-0.0814	0.0084		-0.0224	0.0055	-0.0168	0.0048	-0.0099	0.0055
100		-0.44	-0.1115	0.0202	-0.1223	0.0247	-0.0977	0.0186	-0.64	-0.0763	0.0212	-0.0642	0.0184	-0.0523	0.0142
500			-0.0526	0.0051	-0.0542	0.0049	-0.0444	0.0039		-0.0285	0.0066	-0.0173	0.0055	-0.0119	0.0059
1000			-0.0419	0.0025	-0.0399	0.0026	-0.0325	0.0021		-0.0155	0.0029	-0.0116	0.0022	-0.0094	0.0018
100		-0.3	-0.0461	0.0168	-0.0835	0.0165	-0.0734	0.0143	-0.3	-0.0365	0.0128	-0.0250	0.0110	-0.0186	0.0099
500			-0.0233	0.0030	-0.0314	0.0027	-0.0275	0.0024		-0.0125	0.0031	-0.0127	0.0023	-0.0092	0.0020
1000			-0.0170	0.0016	-0.0209	0.0013	-0.0184	0.0011		-0.0092	0.0013	-0.0072	0.0011	-0.0045	0.0009
100	0.2	-0.7	-0.2850	0.0970	-0.2765	0.0877	-0.2527	0.0739	-0.7	-0.1567	0.0547	-0.1478	0.0501	-0.1216	0.0564
500			-0.1509	0.0398	-0.1425	0.0235	-0.1363	0.0222		-0.0794	0.0222	-0.0650	0.0155	-0.0631	0.0146
1000			-0.1326	0.0222	-0.1210	0.0156	-0.1121	0.0145		-0.0637	0.0096	-0.0562	0.0082	-0.0516	0.0080
100		-0.44	-0.1728	0.0528	-0.1677	0.0462	-0.1527	0.0449	-0.64	-0.1245	0.0430	-0.1171	0.0381	-0.1037	0.0313
500			-0.0748	0.0150	-0.0605	0.0051	-0.0582	0.0049		-0.0510	0.0087	-0.0463	0.0072	-0.0474	0.0068
1000			-0.0433	0.0055	-0.0425	0.0025	-0.0389	0.0024		-0.0319	0.0035	-0.0291	0.0015	-0.0289	0.0015
100		-0.3	-0.1242	0.0188	-0.1168	0.0370	-0.1018	0.0191	-0.3	-0.0585	0.0218	-0.0473	0.0161	-0.0382	0.0153
500			-0.0462	0.0037	-0.0357	0.0026	-0.0300	0.0026		-0.0149	0.0029	0.0085	0.0019	-0.0082	0.0019
1000			-0.0244	0.0018	-0.0166	0.0011	-0.0153	0.0011		-0.0085	0.0012	0.0058	0.0008	-0.0030	0.0008
100	0.4	-0.7	-0.2729	0.0927	-0.2667	0.0838	-0.2597	0.0781	-0.7	-0.2078	0.0819	-0.1992	0.0732	-0.2067	0.0643
500			-0.1462	0.0258	-0.1382	0.0244	-0.1312	0.0199		-0.1321	0.0363	-0.1246	0.0283	-0.1106	0.0278
1000			-0.1166	0.0194	-0.0945	0.0162	-0.1047	0.0155		-0.0832	0.0297	-0.0712	0.0214	-0.0664	0.0201
100		-0.44	-0.1828	0.0578	-0.1750	0.0479	-0.1734	0.0393	-0.64	-0.1141	0.0453	-0.1064	0.0362	-0.1261	0.0324
500			-0.0791	0.0148	-0.0692	0.0108	-0.0585	0.0099		-0.0491	0.0096	-0.0349	0.0071	-0.0388	0.0052
1000			-0.0567	0.0057	-0.0358	0.0056	-0.0285	0.0049		-0.0282	0.0039	-0.0163	0.0016	-0.0202	0.0022
100		-0.3	-0.0903	0.0192	-0.0892	0.0182	-0.0922	0.0183	-0.3	-0.0542	0.0262	-0.0513	0.0172	-0.0583	0.0167
500			-0.0385	0.0028	-0.0325	0.0027	-0.0333	0.0029		-0.0257	0.0054	0.0172	0.0038	-0.0152	0.0036
1000			-0.0232	0.0012	-0.0129	0.0009	-0.0177	0.0010		-0.0178	0.0026	0.0101	0.0018	-0.0090	0.0018

Table 6 Estimates of the bias and MSE of $\widehat{d}_1$ for the FML, Whittle, WI, TML and CSS estimators corresponding to TDGP: $ARFIMA(0, d_0, 0)$ $d_0 = 0.2$, $d^* = 0.0$. The unknown mean is estimated using the sample mean.

	Exact Whittle		TML		CSS	
n	Bias	MSE	Bias	MSE	Bias	MSE
	Correct $ARFIMA(0, d, 0)$ model					
100	-0.0639	0.0184	-0.0542	0.0149	-0.0585	0.0158
500	-0.0163	0.0036	-0.0111	0.0027	-0.0156	0.0032
1000	-0.0052	0.0018	-0.0047	0.0006	-0.0048	0.0006
	Over-Parameterized $ARFIMA(1, d, 0)$ model					
100	-0.0825	0.0282	-0.0758	0.0224	-0.0701	0.0022
500	-0.0267	0.0081	-0.0195	0.0077	-0.0188	0.0075
1000	-0.0102	0.0031	-0.0087	0.0029	-0.0085	0.0029

6 Discussion

This paper presents theoretical and simulation-based results relating to the estimation of mis-specified models for fractionally integrated processes. We show that under mis-specification four classical parametric estimation methods, frequency domain maximum likelihood (FML), Whittle, time domain maximum likelihood (TML) and conditional sum of squares (CSS) converge to the same pseudo-true parameter value. A general closed-form solution for the limiting criterion function for the four alternative estimators is derived in the case of ARFIMA models. This enables us to demonstrate the link between any form of mis-specification of the short memory dynamics and the difference between the true and pseudo-true values of the fractional index, d, and, hence, to the resulting (asymptotic) distributional properties of the estimators, having proved that the estimators are asymptotically equivalent. Consistency of the four estimators for the pseudo-true value is proved for fractional exponents of both the true and estimated models in the long memory, short memory and antipersistent ranges.

The finite sample performance of all four estimators is then documented. The extent to which the finite sample distributions mimic the (numerically specified) asymptotic distributions is displayed. In the case of more extreme mis-specification, and conditional on the mean of the process being known, the pairs of time domain and frequency domain estimators tend to cluster together for smaller sample sizes, with the former pair mimicking the asymptotic distributions more closely. Further, bias and mean squared error (MSE) calculations demonstrate the superiority overall of the CSS estimator, under mis-specification, and the distinct inferiority of the FML estimator – as estimators of the pseudo-true parameter for which they are both consistent. Numerical results for the time-domain estimators in the case where the unknown mean is estimated tell a different story, however, with the Whittle estimator being the superior finite sample performer overall. Numerical results presented for an exact version of the Whittle estimator show a slight superiority over the (approximate) version of the Whittle procedure, in the case where the mean is known; however, the overall ranking of the two methods is reversed when the mean is estimated, with the exact Whittle method not sharing the mean-invariance property of its inexact counterpart.

Note also that very similar results are obtained if the sample mean is replaced by a feasible (plug in) version of the (asymptotically) best linear unbiased estimator (BLUE) of μ.

There are several interesting issues that arise from the results that we have established, including the following: First, Gaussianity underlies the definition of all the estimators considered here, other than CSS perhaps, and Gaussianity is therefore the leading paradigm and an obvious assumption to adopt. Nevertheless, the necessity to suppose that $\{y_t\}$ is a Gaussian process in order to appeal to existing results in the literature where this assumption is made is unfortunate. It seems reasonable to suppose that our results can be extended to fractional linear processes, for example, given that under current assumptions the series will have such a representation. But extension to more general processes is not likely to be straightforward, some idea of the complexities involved can be gleaned from the work of Cavaliere, Nielsen and Taylor (2015), where correctly specified fractional linear processes driven by conditional and unconditional heteroskedastic shocks are examined. Second, although the known (zero) mean assumption is inconsequential for the FML and Whittle estimators, this is not the case for the exact Whittle and time domain estimators, as our bias and mean squared error experimental results obtained using demeaned data show. The deterioration in the overall performance of the exact Whittle and time domain estimators once the estimation of μ plays a role in their computation might have been anticipated since the rate of convergence of the sample average to μ is $n^{1/2-d_0}$ (Hosking, 1996, Theorem 8), and thus slower the larger the value of d_0. Similarly, estimation of μ also seems likely to impact on the limiting distribution of the time domain estimators – because the rate of convergence of the estimators when the true mean is known is $n^{1-2d^*}/\log n$ when $d^* = d_0 - d_1 > 0.25$, $(n/\log\log\log n)^{1/2}$ when $d^* = 0.25$, and $\sqrt{n}$ otherwise – something that we have not investigated here. Third, the extension of our results to non-stationary cases will facilitate the consideration of a broader range of circumstances. To some extent non-stationary values of d might be covered by means of appropriate pre-filtering, for example, the use of first-differencing when $d_0 \in (1/2, 3/2)$, but this would require prior knowledge of the structure of the process and opens up the possibility of a different type of mis-specification from the one we have considered here. Explicit consideration of the non-stationary case with $d \in (0, 3/2)$, say, perhaps offers a better approach as prior knowledge of the characteristics of the process would then be unnecessary. The latter also seems particularly relevant given that estimates near the boundaries $d = 0.5$ and $d = 1$ are not uncommon in practice. Previous developments in the analysis of non-stationary fractional processes (see, inter alios, Beran, 1995; Tanaka, 1999; Velasco, 1999; Velasco and Robinson, 2000) might offer a sensible starting point for such an investigation. Fourth, our limiting distribution results can be used in practice to conduct inference on the long memory and other parameters after constructing obvious smoothed periodogram consistent estimates of $\mathbf{B}$, $\boldsymbol{\mu}_n$, $\overline{\Lambda}_{dd}$ and $\mathbf{\Lambda}$. But which situation should be assumed in any particular instance, $d^* > 0.25$, $d^* = 0.25$ or $d^* < 0.25$, may be a moot point. Fifth, the relationships between the bias and MSE of the parametric estimators of d_1 (denoted respectively below by Bias_d_1 and MSE_d_1), and the bias and MSE as estimators of the *true* value d_0, (Bias_d_0 and MSE_d_0 respectively) can be expressed simply as follows:

$$\begin{aligned}
\text{Bias_}d_0 &= E_0(\widehat{d}_1) - d_0 \\
&= \left[E_0(\widehat{d}_1) - d_1\right] + (d_1 - d_0) \\
&= Bias_d_1 - d^*\,,
\end{aligned}$$

where we recall, $d^* = d_0 - d_1$, and

$$\begin{aligned}
\text{MSE_}d_0 &= E_0\left(\widehat{d}_1 - d_0\right)^2 \\
&= E_0\left(\widehat{d}_1 - E_0(\widehat{d}_1)\right)^2 + \left[E_0(\widehat{d}_1) - d_0\right]^2 \\
&= E_0\left(\widehat{d}_1 - E_0(\widehat{d}_1)\right)^2 + \left[[E_0(\widehat{d}_1) - d_1] - d^*\right]^2 \\
&= E_0\left(\widehat{d}_1 - E_0(\widehat{d}_1)\right)^2 + \left[E_0(\widehat{d}_1) - d_1\right]^2 + d^{*2} - 2d^*\left[E_0(\widehat{d}_1) - d_1\right] \\
&= MSE_d_1 + d^{*2} - 2d^* Bias_d_1.
\end{aligned}$$

Hence, if Bias_d_1 is the same sign as d^* at any particular point in the parameter space, then the bias of a mis-specified parametric estimator *as an estimator of* d_0, may be less (in absolute value) than its bias as an estimator of d_1, depending on the magnitude of the two quantities. Similarly, MSE_d_0 may be less than MSE_d_1 if Bias_d_1 and d^* have the same sign, with the final result again depending on the magnitude of the two quantities. These results imply that it is possible for the ranking of mis-specified parametric estimators to be altered, once the reference point changes from d_1 to d_0. This raises the following questions: Does the dominance of the CSS estimator (within the parametric set of estimators) – and in the known mean case – still obtain when the true value of d is the reference value? And more critically from a practical perspective; Are there circumstances where a mis-specified parametric estimator out-performs semi-parametric alternatives in finite samples, the lack of consistency (for d_0) of the former notwithstanding? Such topics remain the focus of current and ongoing research.

References

Beran, J. (1994). *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society* **B 57** 654–672.

Boes, D. C., Davis, R. A. and Gupta, S. N. (2006). Parameter estimation in low order fractionally differenced ARMA processes. *Stochastic Hydrology and Hydraulics* **3** 97–110.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer-Verlag, New York, 2nd ed.

Cavaliere, G., Nielsen, M. O. and Taylor, A. M. R. (2015). Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form. *Queens Economics Department Working Paper* No. 1324, Queen's University, Ontario.

Chen, W. W. and Deo, R. S. (2006). Estimation of mis-specified long memory models. *Journal of Econometrics* **53** 257 – 281.

Cheung, Y. W. and Diebold, F. X. (1994). On maximum likelihood estmation of the differencing parameter of fractionally integrated noise with unknown mean. *Journal of Econometrics* **62** 301–316.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics* **17** 1749–1766.

Doornik, J. A. and Ooms, M. (2001). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis* **42** 333–348. Also a 2001 Nuffield discussion paper.

Fox, R. and Taqqu, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Annals of Statistics* **14** 517–532.

Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Tables of Integrals, Series and Products.* Academic Press, Sydney.

Grenander, U. and Szego, G. (1958). *Toeplitz Forms and Their Application.* University of California Press, Berkeley.

Hannan, E. J. (1973). The asymptotic theory of linear time series models. *Advances in Applied Probability* **10** 130–145.

Hurvich, C. M. and Beltrao, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long memory time series. *Journal of Time Series Analysis* **14** 455–472.

Hosking, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long memory time series. *Journal of Econometrics* **73** 261–284.

Lieberman, O. and Phillips, P. C. B. (2004). Error bounds and asymptotic expansions for toeplitz product functionals of unbounded spectra. *Journal of Time Series Analysis* **25** 732–753.

Nielsen, M. O. and Frederiksen, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* **24** 405–443.

Robinson, P. M. (2006). *Time Series and Related Topics: In Memory of Ching-Zong Wei*, vol. 52 of *IMS Lecture Notes–Monograph Series*, chap. Conditional-Sum-of-Squares Estimation of Models for Stationary Time Series with Long Memory. Institue of Mathematical Statistics, Beachwood, 130–137.

Sowell, F. (1992). Maximum likelihood estmation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* **53** 165–188.

Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory* **15** 549–582.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20** 87–127.

Velasco, C. and Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* **95:452** 1229–1243.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.

Whittle, P. (1984). *Prediction and Regulation By Linear Least Squares Methods.* Basil Blackwell.

Yajima, Y. (1992). Asymptotic properties of estimates in incorrect ARMA models for long-memory time series. In D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. Taqqu, eds., *New Directions in Time Series Analysis, Part II*, vol. 46 of *IMA Volumes in Mathematics and Its Applications.* Springer-Verlag, New York, 375–382.

A Proofs

In the proofs we will need to consider stochastic Rieman-Stieltges integrals of the periodogram. These are dealt with in the following lemma.

Lemma A.1 *Assume that $I(\lambda)$ is calculated from a realization of a stationary Gaussian process with a spectral density as given in* (1)*, and that $h(\lambda)$ is an even valued periodic function with period 2π that is continuously differentiable on $(0,\pi]$. Set*

$$\nabla_I(h) = \int_0^\pi I(\lambda)h(\lambda)d\lambda - \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor} I(\lambda_j)h(\lambda_j)\,.$$

Then $\nabla_I(h) = O_p(n^{-1})$ and $\lim_{n\to\infty}|\nabla_I(h)| = 0$ almost surely.

Proof. Using the partition of $(0,\pi]$ induced by $\lambda_j = 2\pi j/n$, $j = 1,\ldots,\lfloor n/2\rfloor$, gives the decomposition

$$\nabla_I(h) = \int_0^{2\pi/n} I(\lambda)h(\lambda)d\lambda + \sum_{j=1}^{\lfloor n/2\rfloor-1}\int_{2\pi j/n}^{2\pi(j+1)/n}\{I(\lambda)h(\lambda) - I(\lambda_j)h(\lambda_j)\}d\lambda$$
$$+\int_{2\pi\lfloor n/2\rfloor/n}^{\pi} I(\lambda)h(\lambda)d\lambda - I(\lambda_{\lfloor n/2\rfloor})h(\lambda_{\lfloor n/2\rfloor})\frac{2\pi}{n}\,,$$

which can be rearranged to give $\nabla_I(h) = T_1 + T_2 + T_3$ where $T_1 = \int_0^{2\pi/n} I(\lambda)h(\lambda)d\lambda$,

$$T_2 = \sum_{j=1}^{\lfloor n/2\rfloor-1}\int_{2\pi j/n}^{2\pi(j+1)/n} I(\lambda)\{h(\lambda) - h(\lambda_j)\}d\lambda + \int_{2\pi\lfloor n/2\rfloor/n}^{\pi} I(\lambda)\{h(\lambda) - h(\lambda_{\lfloor n/2\rfloor})\}d\lambda$$

and

$$T_3 = \sum_{j=1}^{\lfloor n/2\rfloor-1}\int_{2\pi j/n}^{2\pi(j+1)/n}\{I(\lambda)-I(\lambda_j)\}h(\lambda_j)d\lambda + \int_{2\pi\lfloor n/2\rfloor/n}^{\pi} I(\lambda)h(\lambda_{\lfloor n/2\rfloor})d\lambda - I(\lambda_{\lfloor n/2\rfloor})h(\lambda_{\lfloor n/2\rfloor})\frac{2\pi}{n}\,.$$

By the first mean value theorem for integrals $T_1 = I(\lambda')h(\lambda')\frac{2\pi}{n}$ where $\lambda' \in (0,\lambda_1]$. Set $(2\pi n)^{-1/2}\sum_{t=1}^n y_t\exp(-i\lambda t) = A(\lambda) - iB(\lambda)$ and let $\delta > 0$ denote an arbitrary constant. Since $\lim_{(\lambda_1-\lambda')\to 0}|A(\lambda') - A(\lambda_1)| = 0$ and $\lim_{(\lambda_1-\lambda')\to 0}|B(\lambda') - B(\lambda_1)| = 0$ it follows that $|A(\lambda') - A(\lambda_1)| < \delta$ and $|B(\lambda') - B(\lambda_1)| < \delta$ for all n sufficiently large, and by Theorem 3 of Hurvich and Beltrao (1993) the normalised Fourier coefficients $A(\lambda_1)/\{f_0(\lambda_1)\}^{1/2}$ and $B(\lambda_1)/\{f_0(\lambda_1)\}^{1/2}$ are asymptotically independent Gaussian random variables. From the inequality

$$|T_1| \le (A^2(\lambda') + B^2(\lambda'))\cdot|h(\lambda')|\frac{2\pi}{n}\,.$$

it therefore follows that $|T_1|$ is $O_p(n^{-1})$. For the second term we have

$$|T_2| \le \sum_{j=1}^{\lfloor n/2\rfloor}\int_{2\pi j/n}^{2\pi(j+1)/n} I(\lambda)|h(\lambda) - h(\lambda_j)|d\lambda\,.$$

But for all $\lambda \in (\lambda_j,\lambda_{j+1})$, $|h(\lambda) - h(\lambda_j)| \le \mathcal{M}\frac{2\pi}{n}$ where $\mathcal{M} = \sup_{\lambda\in[\frac{2\pi}{n},\pi]}|dh(\lambda)/d\lambda|$, and it follows that $|T_2| \le \int_{-\pi}^{\pi} I(\lambda)d\lambda\mathcal{M}\pi/n = \pi\mathcal{M}\sum_{t=1}^n y_t^2/n^2$. Since $\sum_{t=1}^n y_t^2/n$ converges to $E_0(y_t^2)$ (Hosking, 1996, Theorem 4), we can conclude that $|T_2|$ is $O_p(n^{-1})$. Similarly,

$$|T_3| \le \sum_{j=1}^{\lfloor n/2\rfloor} I(\lambda_j)\int_{2\pi j/n}^{2\pi(j+1)/n}\left|\frac{I(\lambda)}{I(\lambda_j)} - 1\right||h(\lambda_j)|d\lambda\,,$$

and if $|I(\lambda) - I(\lambda_j)| < I(\lambda_j)$ for all $\lambda \in (\lambda_j, \lambda_{j+1})$, $j = 1, \ldots, \lfloor n/2 \rfloor$, then it follows that $|T_3| \leq \sum_{j=1}^{\lfloor n/2 \rfloor} I(\lambda_j)\mathcal{M}' 2\pi/n = 2\pi\mathcal{M}' \sum_{t=1}^{n} y_t^2/n^2$ where $\mathcal{M}' = \sup_{\lambda \in [\frac{2\pi}{n}, \pi]} |h(\lambda)|$. Now the event $\sup_{\lambda \in (\lambda_j, \lambda_{j+1}]} |I(\lambda) - I(\lambda_j)| < I(\lambda_j)$ is equivalent to $0 < \sup_{\lambda \in (\lambda_j, \lambda_{j+1}]} I(\lambda) < 2I(\lambda_j)$ and hence

$$Pr\{\sup_{\lambda \in (\lambda_j, \lambda_{j+1}]} |I(\lambda) - I(\lambda_j)| < I(\lambda_j)\} = 1 - Pr\{2I(\lambda_j) \leq \sup_{\lambda \in (\lambda_j, \lambda_{j+1}]} I(\lambda)\}.$$

Moreover, since $(\lambda_{j+1} - \lambda_j) = 2\pi/n \to 0$ as $n \to \infty$ and $\lim_{(\lambda - \lambda_j) \to 0} |I(\lambda) - I(\lambda_j)| = 0$ it follows that for all n sufficiently large $|\sup_{\lambda \in (\lambda_j, \lambda_{j+1}]} I(\lambda) - I(\lambda_j)| < \delta$ and

$$\begin{aligned}
\lim_{n \to \infty} Pr\{\sup_{\lambda \in (\lambda_j, \lambda_{j+1}]} |I(\lambda) - I(\lambda_j)| < I(\lambda_j)\} &= 1 - \lim_{n \to \infty} Pr\{2I(\lambda_j) \leq \sup_{\lambda \in (\lambda_j, \lambda_{j+1}]} I(\lambda)\} \\
&\geq 1 - \lim_{n \to \infty} Pr\{I(\lambda_j) \leq 0.5\{I(\lambda_j) + \delta\}\} \\
&= 1 - \lim_{n \to \infty} Pr\{I(\lambda_j) \leq \delta\}.
\end{aligned}$$

Because δ is arbitrary, and the normalised periodogram $I(\lambda_j)/f_0(\lambda_j)$ converges in distribution to a weighted sum of independent $\chi^2(1)$ variables (Hurvich and Beltrao, 1993, Theorem 6), it follows that $|T_3| \leq 2\pi\mathcal{M}' \sum_{t=1}^{n} y_t^2/n^2$ with probability approaching one as $n \to \infty$ and $|T_3|$ is therefore $O_p(n^{-1})$. We have thus established that $\nabla_I(h)$ is bounded above by three terms each of order $O_p(n^{-1})$. Furthermore, since each term has a variance of order $O(n^{-2})$ it follows from Markov's inequality and the Borel-Cantelli lemma that $\nabla_I(h)$ converges to zero almost surely. ■

A.1 Proof of Lemmas 1 and 2

A.1.1 Proof of Lemma 1:

Let $\mathbf{A}_\eta = [a_{s-r}(\eta)]$ where

$$a_{s-r}(\eta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_1(\boldsymbol{\eta}, \lambda)} \exp(i(s-l)\lambda) d\lambda, \quad r, s = 1, \ldots, n. \tag{A.1}$$

Then

$$\begin{aligned}
\frac{1}{n}\mathbf{Y}^T \mathbf{A}_\eta \mathbf{Y} &= \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{f_1(\boldsymbol{\eta}, \lambda)} \exp(ik\lambda) d\lambda \sum_{t=k}^{n} y_t y_{t-k} \\
&= \int_{-\pi}^{\pi} \frac{1}{f_1(\boldsymbol{\eta}, \lambda)} \left(\frac{1}{2\pi n} \sum_{k=0}^{n-1} \exp(ik\lambda) \sum_{t=k}^{n} y_t y_{t-k} \right) d\lambda \\
&= \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda,
\end{aligned} \tag{A.2}$$

and by Lemma A.1 it follows that

$$\left| \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} - \frac{\mathbf{Y}^T \mathbf{A}_\eta \mathbf{Y}}{2n} \right| = \left| \frac{2\pi}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} - \int_0^{\pi} \frac{I(\lambda)}{f_1(\boldsymbol{\eta}, \lambda)} d\lambda \right| = O_p(n^{-1}).$$

Hence it follows that, up to terms of order n^{-1}, the behaviour of $Q_n^1(\boldsymbol{\eta})$ as $n \to \infty$ can be ascertained by examining that of $\frac{1}{n}\mathbf{Y}^T \mathbf{A}_\eta \mathbf{Y}$. Now, $E_0[\mathbf{Y}^T \mathbf{A}_\eta \mathbf{Y}] = \mathrm{tr}\mathbf{A}_\eta \boldsymbol{\Sigma}_0 = \mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0 +$

$\mathrm{tr}(\mathbf{A}_\eta - \mathbf{\Sigma}_\eta^{-1})\mathbf{\Sigma}_0$, where $\mathbf{\Sigma}_0 = E_0[\mathbf{Y}\mathbf{Y}^T]$, and

$$\begin{aligned}
\mathrm{tr}\mathbf{\Sigma}_\eta^{-1}\mathbf{\Sigma}_0 + \mathrm{tr}(\mathbf{A}_\eta - \mathbf{\Sigma}_\eta^{-1})\mathbf{\Sigma}_0 &= \|\mathbf{\Sigma}_\eta^{-\frac{1}{2}}\mathbf{\Sigma}_0^{\frac{1}{2}}\|^2 + \|(\mathbf{A}_\eta - \mathbf{\Sigma}_\eta^{-1})^{\frac{1}{2}}\mathbf{\Sigma}_0^{\frac{1}{2}}\|^2 \\
&\geq \|\mathbf{\Sigma}_\eta^{-\frac{1}{2}}\mathbf{\Sigma}_0^{\frac{1}{2}}\|^2_{spec} \\
&= O(n^{2(d_0-d)\wedge 0})
\end{aligned}$$

where $\|\mathbf{A}\|^2_{spec} = \sup_{\|x\|=1} \mathbf{x}'\mathbf{A}\mathbf{A}'\mathbf{x}$. The last line follows from an adaptation of the proof of Lemma 5.3 of (Dahlhaus, 1989, page 1761) that shows that Dahlhaus' Lemma 5.3 holds true for negative exponents. Since the variance of $Q_n^1(\boldsymbol{\eta})$ is $O(n^{-1})$ we can therefore conclude from a one-sided Chebychev inequality that $Pr\left(Q_n^1(\boldsymbol{\eta}) \leq n^{-1}\mathrm{tr}\mathbf{A}_\eta\mathbf{\Sigma}_0 - \epsilon\right)$ can be made arbitrarily small for all $\epsilon > 0$ by taking n sufficiently large, and hence that $Q_n^1(\boldsymbol{\eta}) = O_p(n^{2(d_0-d)\wedge 0-1})$. This establishes the first part of the lemma. The proof of the second part of the lemma follows an argument first developed by Hannan (1973) in the context of short memory processes. This argument can be extended to fractional processes, see the comments in (Brockwell and Davis, 1991, pages 472-473). The details, which follow the steps used by (Brockwell and Davis, 1991, Section 10.8.2, pages 378-379) in the proof of their Proposition 10.8.2, are omitted.

A.1.2 Proof of Lemma 2:

Since by construction $h_1(\boldsymbol{\eta}, \lambda) > 0$ the result can be proved by once again employing the argument used by (Brockwell and Davis, 1991, Section 10.8.2, pages 378-379) in their proof of their Proposition 10.8.2, the detailed steps are omitted.

A.2 Proof of Propositions 1 and 2

A.2.1 Proof of Proposition 1:

It only remains for us to establish that $Q(\boldsymbol{\eta})$ provides a limit superior for $Q_n^{(1)}(\boldsymbol{\eta})$ when $(d_0 - d) < 0.5$ and $d < 0$. In this case $f_1(\boldsymbol{\eta}, \lambda) = |\lambda|^{2|d|}L(\lambda)$ where $L(\lambda)$ is slowly varying and bounded as $\lambda \to 0$ and there exists an $\epsilon \in (0, 2|d|)$ and a $K > 0$, that may depend on ϵ, such that $f_1(\boldsymbol{\eta}, \lambda) = |\lambda|^{2|d|}K|\lambda|^{-\epsilon}$. We therefore have that $f_1(\boldsymbol{\eta}, \lambda) > K|\lambda|^{2|d|}$ when $|\lambda| < 1$ and $h_1(\boldsymbol{\eta}, \lambda) \neq f_1(\boldsymbol{\eta}, \lambda)$ whenever $\lambda < (K^{-1}\delta)^{1/(2|d|-\epsilon)}$, from which it follows that

$$Q_n^{(1)}(\boldsymbol{\eta}) \leq \frac{2\pi}{n}\sum_{j=k_\delta+1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} + \frac{1}{K}\left(\frac{2\pi}{n}\right)^{1-2|d|}\sum_{j=1}^{k_\delta} I(\lambda_j) \tag{A.3}$$

where $k_\delta = \lfloor (K^{-1}\delta)^{1/(2|d|-\epsilon)}(2\pi/n)\rfloor + 1$. The inequality in (A.3) follows because for all $\lambda_j < (K^{-1}\delta)^{1/(2|d|-\epsilon)} < 2\pi k_\delta/n$ we have

$$\left(\frac{h_1(\boldsymbol{\eta}, \lambda_j)}{f_1(\boldsymbol{\eta}, \lambda_j)} - 1\right) \leq \left(\frac{\delta}{K}\left(\frac{n}{2\pi}\right)^{2|d|} - 1\right)$$

and

$$\begin{aligned}
Q_n^{(1)}(\boldsymbol{\eta}) &= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} + \frac{2\pi}{n}\sum_{j=1}^{k_\delta} I(\lambda_j)\left(\frac{1}{f_1(\boldsymbol{\eta}, \lambda_j)} - \frac{1}{h_1(\boldsymbol{\eta}, \lambda_j)}\right) \\
&\leq \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}_1, \lambda_j)} + \frac{2\pi}{n}\sum_{j=1}^{k_\delta}\frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)}\left(\frac{\delta}{K}\left(\frac{n}{2\pi}\right)^{2|d|} - 1\right) \\
&= \frac{2\pi}{n}\sum_{j=k_\delta+1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{h_1(\boldsymbol{\eta}, \lambda_j)} + \frac{1}{K}\left(\frac{2\pi}{n}\right)^{1-2|d|}\sum_{j=1}^{k_\delta} I(\lambda_j)\,.
\end{aligned}$$

Applying Lemma 2 to the first term on the right hand side in (A.3) gives a limit of

$$\frac{\sigma_0^2}{2\pi}\int_{(K^{-1}\delta)^{1/(2|d|-\epsilon)}}^{\pi}\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta}_1,\lambda)}d\lambda\,.$$

Similarly

$$\lim_{n\to\infty}\frac{2\pi}{n}\sum_{j=1}^{k_\delta}I(\lambda_j)=\frac{\sigma_0^2}{2\pi}\int_0^{(K^{-1}\delta)^{1/(2|d|-\epsilon)}}f_0(\lambda)d\lambda=\frac{\sigma_0^2}{2\pi}f_0(\lambda'^{-1}\delta)^{1/(2|d|-\epsilon)}$$

for some $\lambda'^{-1}\delta)^{1/(2|d|-\epsilon)})$ by the first mean value theorem for integrals. Setting $\delta=(2\pi)^{2|d|-\epsilon}/n^p$ where $p>2|d|-\epsilon$, we find that

$$\begin{aligned}\frac{1}{K}\left(\frac{2\pi}{n}\right)^{1-2|d|}\sum_{j=1}^{k_\delta}I(\lambda_j) &\sim \frac{1}{K}\left(\frac{n}{2\pi}\right)^{2|d|}\frac{\sigma_0^2}{2\pi}f_0(\lambda')\frac{2\pi k_\delta}{n}\\ &\sim \frac{1}{K}\left(\frac{2\pi}{n}\right)^{1-2|d|}\frac{\sigma_0^2}{2\pi}f_0(\lambda')\left(\frac{1}{n}\right)^{\frac{p-2|d|+\epsilon}{(2|d|-\epsilon)}}\end{aligned}$$

and hence we can conclude that

$$\limsup_{n\to\infty}Q_n^{(1)}(\boldsymbol{\eta})\le Q(\boldsymbol{\eta})$$

uniformly in $\boldsymbol{\eta}\in\mathbb{E}$ when $(d_0-d)<0.5$ and $d<0$, as required.

A.2.2 Proof of Proposition 2:

Let $\boldsymbol{\eta}_n$ denote a sequence in $\mathbb{E}$ that converges to $\boldsymbol{\eta}$. From the assumed continuity of $f_1(\boldsymbol{\eta},\lambda)$, for any $\delta>0$ we can determine a value n' such that $|f_1(\boldsymbol{\eta}_n,\lambda)-f_1(\boldsymbol{\eta},\lambda)|<\delta^3$ for all $n>n'$ and

$$\begin{aligned}\left|\frac{1}{f_1(\boldsymbol{\eta}_n,\lambda)+\delta}-\frac{1}{f_1(\boldsymbol{\eta},\lambda)+\delta}\right| &= \left|\frac{|f_1(\boldsymbol{\eta}_n,\lambda)-f_1(\boldsymbol{\eta},\lambda)|}{(f_1(\boldsymbol{\eta}_n,\lambda)+\delta)(f_1(\boldsymbol{\eta},\lambda)+\delta)}\right|\\ &\le \frac{|f_1(\boldsymbol{\eta}_n,\lambda)-f_1(\boldsymbol{\eta},\lambda)|}{\delta^2}\\ &= \delta\end{aligned}$$

uniformly in $[-\pi,\pi]$. Consequently

$$\left|\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}_n,\lambda_j)+\delta}-\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)+\delta}\right|\le\frac{2\delta\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}I(\lambda_j)\tag{A.4}$$

and, using the second part of Lemma 1 in conjunction with (A.4), it follows that

$$\begin{aligned}\liminf_{n\to\infty}Q_n^{(1)}(\boldsymbol{\eta}_n) &\ge \liminf_{n\to\infty}\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta}_n,\lambda_j)+\delta}\\ &\ge \lim_{n\to\infty}\left\{\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)+\delta}-\frac{2\delta\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}I(\lambda_j)\right\}\\ &= \frac{\sigma_0^2}{2\pi}\int_0^{\pi}\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)+\delta}d\lambda-\delta\pi\gamma_0(0)\,,\end{aligned}$$

where $\gamma_0(0)$ is the variance of the TDGP. Letting $\delta \to 0$ and applying Lebegue's monotone convergence theorem gives

$$\liminf_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}_n) \geq \frac{\sigma_0^2}{2\pi}\int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda = Q(\boldsymbol{\eta})\,.$$

Since by definition $\boldsymbol{\eta}_1$ minimises $Q(\boldsymbol{\eta})$ it follows that $Q(\boldsymbol{\eta}_1)$ provides a lower bound to the limit inferior of any sequence in $\mathbb{E}$. By definition of $\widehat{\boldsymbol{\eta}}_1^{(1)}$, however, it follows from Proposition 1 that

$$\limsup_{n\to\infty} Q_n^{(1)}(\widehat{\boldsymbol{\eta}}_1^{(1)}) \leq \limsup_{n\to\infty} Q_n^{(1)}(\boldsymbol{\eta}_1) = Q(\boldsymbol{\eta}_1)\,.$$

We can therefore conclude that $|Q_n^{(1)}(\boldsymbol{\eta}_1^{(1)}) - Q(\boldsymbol{\eta}_1)| \to 0$ almost surely and an argument by contradiction then shows that $\widehat{\boldsymbol{\eta}}_1^{(1)} \to \boldsymbol{\eta}_1$ with probability one.

A.3 Proof of Theorem 1:

In what follows we assume that the mean is known, and without loss of generality set $\mu = 0$.

A.3.1 The Whittle estimator

Concentrating $Q_n^{(2)}(\sigma^2,\boldsymbol{\eta})$ with respect to σ^2 and setting $n\lfloor n/2\rfloor = 0.5$ yields the profile (negative) log-likelihood

$$Q_n^{(2)}(\boldsymbol{\eta}) = \frac{2\pi}{2}\log\left(\frac{\widehat{\sigma}^2(\boldsymbol{\eta})}{2\pi}\right) + \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\log f_1(\boldsymbol{\eta},\lambda_j) + \pi$$

where $\widehat{\sigma}^2(\boldsymbol{\eta}) = 2Q_n^{(1)}(\boldsymbol{\eta})$ and $Q_n^{(1)}(\boldsymbol{\eta})$ is as given in (4)). This indicates that the value of $\boldsymbol{\eta}$ that minimises the profile log-likelihood, *having first deleted the term* $2\pi\sum_{j=1}^{\lfloor n/2\rfloor}\log f_1(\boldsymbol{\eta},\lambda_j)/n$, is $\arg\min_{\boldsymbol{\eta}} Q_n^{(1)}(\boldsymbol{\eta})$. Following the development in Beran (1994, p. 116), however, we have

$$\lim_{n\to\infty}\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\log f_1(\boldsymbol{\eta},\lambda_j) = \frac{1}{2}\int_{-\pi}^{\pi}\log f_1(\boldsymbol{\eta},\lambda)d\lambda,$$

where

$$\begin{aligned}\int_{-\pi}^{\pi}\log f_1(\boldsymbol{\eta},\lambda)d\lambda &= \int_{-\pi}^{\pi}\log\left(g_1(\boldsymbol{\beta},\lambda)|2\sin(\lambda/2)|^{-2d}\right)d\lambda \\ &= \int_{-\pi}^{\pi}\log g_1(\boldsymbol{\beta},\lambda)d\lambda - 2d\int_{-\pi}^{\pi}\log|2\sin(\lambda/2)|d\lambda\,.\end{aligned}$$

By Assumption A.6 $M : ARFIMA$(0,d,0). $\int_{-\pi}^{\pi}\log g_1(\boldsymbol{\beta},\lambda)d\lambda = 0$, and from standard results for trigonometric integrals (Gradshteyn and Ryzhik, 2007, p.583) we have $\int_{-\pi}^{\pi}\log|2\sin(\lambda/2)|d\lambda = 2\int_0^{\pi}\log|2\sin(\lambda/2)|d\lambda = 0$, The asymptotic equivalence of the FML and Whittle estimators then follows.

A.3.2 The TML estimator

From the triangular inequality we have

$$\left|\int_0^\pi \frac{I(\lambda)-f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda\right| \le \left|\int_0^\pi \frac{I(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda - \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\right| + \left|\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)} - \int_0^\pi \frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda\right|$$

and it follows by Lemma A.1 and application of Proposition 1 in conjunction with (A.2) that

$$\text{plim}_{n\to\infty}\left|\frac{1}{n}\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y} - \int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda\right| = 0\,. \tag{A.5}$$

Now, $E_0[\mathbf{Y}^T(\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\mathbf{Y}] = \text{tr}(\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0 = E_0[\mathbf{Y}\mathbf{Y}^T]$, and $|\text{tr}(\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\boldsymbol{\Sigma}_0| \le \|\mathbf{I}-\boldsymbol{\Sigma}_\eta^{1/2}\mathbf{A}_\eta\boldsymbol{\Sigma}_\eta^{1/2}\|\cdot\|\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0\|_{spec}$. By Lemma 5.2 of Dahlhaus (1989) applied with the function f of Dahlhaus equal to $f_1(\boldsymbol{\eta},\lambda)$ when $0\le d<0.5$ and $f^{-1}=4\pi^2 f_1(\boldsymbol{\eta},\lambda)$ when $-0.5<d<0$, the factor $\|\mathbf{I}-\boldsymbol{\Sigma}_\eta^{1/2}\mathbf{A}_\eta\boldsymbol{\Sigma}_\eta^{1/2}\| = O(n^\delta)$ for all $\delta\in(0,|d|/2)$. By Lemma 5.3 of Dahlhaus (1989) and its proof we have $\|\boldsymbol{\Sigma}_\eta^{-\frac{1}{2}}\boldsymbol{\Sigma}_0^{\frac{1}{2}}\|_{spec}^2 = O(n^{2(d_0-d)\wedge 0})$. It follows that $|\text{tr}(\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\boldsymbol{\Sigma}_0| = O(n^{2(d_0-d)\wedge 0+\delta})$. Similarly $Var_0[\mathbf{Y}^T(\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\mathbf{Y}] = \text{tr}((\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\boldsymbol{\Sigma}_0)^2 \le \|(\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\boldsymbol{\Sigma}_0\|^2 = O(n^{4(d_0-d)\wedge 0+2\delta})$. Markov's inequality therefore implies that

$$Pr\left(n^{-1}\left|\mathbf{Y}^T(\boldsymbol{\Sigma}_\eta^{-1}-\mathbf{A}_\eta)\mathbf{Y}\right|>\epsilon\right) = O(n^{-2+4(d_0-d)\wedge 0+2\delta})$$

for all $\epsilon>0$, and hence $\text{plim}_{n\to\infty}n^{-1}\left|\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}\right| = 0$. Thus we can conclude that

$$\text{plim}_{n\to\infty}\left|\frac{1}{n}\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y} - \int_{-\pi}^{\pi}\frac{f_0(\lambda)}{f_1(\boldsymbol{\eta},\lambda)}d\lambda\right| = 0\,.$$

From Grenander and Szego (1958) we know that

$$\lim_{n\to\infty}\frac{1}{n}\log|\boldsymbol{\Sigma}_\eta| = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_1(\boldsymbol{\eta},\lambda)d\lambda = 0,$$

for the second term in (8). The probability limit of the criterion function $Q_n^{(3)}(\sigma_\varepsilon^2,\boldsymbol{\eta})$ is therefore

$$\mathcal{Q}^{(3)}(\sigma^2,Q(\boldsymbol{\eta})) = \log\sigma^2 + \frac{2Q(\boldsymbol{\eta})}{\sigma^2}$$

uniformly in σ^2 and $\boldsymbol{\eta}$. Concentrating $\mathcal{Q}^{(3)}(\sigma^2,Q(\boldsymbol{\eta}))$ with respect to σ^2 we find that the minimum of $\mathcal{Q}^{(3)}(\sigma^2,Q(\boldsymbol{\eta}))$ is given by $\log(2Q(\boldsymbol{\eta}_1))+1$. Once again $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ is the pseudo-true parameter for the estimator under mis-specification and we conclude that $\text{plim}\,\widehat{\boldsymbol{\eta}}_1^{(3)} = \boldsymbol{\eta}_1$.

A.3.3 The CSS estimator

Let $\mathbf{T}_\eta$ and $hbfH_\eta$ denote the $n\times n$ upper triangular Toeplitz matrix with non-zero elements $\tau_{|i-j|}(\boldsymbol{\eta})$, $i,j=1,\ldots,n$, and the $n\times\infty$ reverse Hankel matrix with typical element $\tau_{n-i+j}(\boldsymbol{\eta})$, $i=1,\ldots,n$, $j=1,\ldots,\infty$, respectively. Then from (A.1) we can deduce that $\mathbf{A}_\eta = \mathbf{T}_\eta\mathbf{T}_\eta^T + \mathbf{H}_\eta\mathbf{H}_\eta^T$. From (10) and (11) it follows that $Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n}\mathbf{Y}^T\mathbf{T}_\eta\mathbf{T}_\eta^T\mathbf{Y}$, and it is shown below that $\frac{1}{n}\mathbf{Y}^T\mathbf{H}_\eta\mathbf{H}_\eta^T\mathbf{Y} = o(1)$ in the relevant region of the parameter space, namely where $d>-1/2$ and $d_0-d<1/2$. We can therefore conclude that $\left|Q_n^{(4)}(\boldsymbol{\eta})-\frac{1}{n}\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}\right|$ converges to zero in probability, and hence, using (A.5), the limiting value of the criterion function $Q_n^{(4)}(\boldsymbol{\eta})$

is $2Q(\boldsymbol{\eta})$. That the pseudo-true parameter for the CSS estimator under mis-specification is $\boldsymbol{\eta}_1 = \arg\min_{\boldsymbol{\eta}} Q(\boldsymbol{\eta})$ and $\operatorname{plim} \widehat{\boldsymbol{\eta}}_1^{(4)} = \boldsymbol{\eta}_1$ is now obvious. It remains for us to establish that $\frac{1}{n}\mathbf{Y}^T\mathbf{H}_\eta\mathbf{H}_\eta^T\mathbf{Y} = o(1)$. Suppressing the dependence on the parameter $\boldsymbol{\eta}$ for notational simplicity, set $\mathbf{M} = \mathbf{H}\mathbf{H}^T$. Then $\mathbf{M} = [m_{ij}]_{i,j=1,\ldots,n}$ where $m_{ij} = \sum_{u=0}^{\infty} \tau_{u+n-i}\tau_{u+n-j}$, and

$$E_0[\mathbf{Y}^T\mathbf{M}\mathbf{Y}] = \operatorname{tr}(\mathbf{M}\boldsymbol{\Sigma}_0) = \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\gamma_0(j-i)$$

where $\gamma_0(\tau)$, $\tau = 0, \pm 1, \pm 2, \ldots$, denotes the autocovariance function of the TDGP. Since $|\tau_k| \sim k^{-(1+d)}\mathcal{C}_\tau$, $\mathcal{C}_\tau < \infty$, the series $\sum_{k=0}^{\infty} |\tau_k|^2 \sim \mathcal{C}_\tau^2\zeta(2(d+1))$ when $d > -1/2$, where $\zeta(\cdot)$ denotes the Riemann zeta function, from which we can deduce that $|m_{ij}| \sim \{(n-i+1)(n-j+1)\}^{-(1+d)}\mathcal{C}_m$, $\mathcal{C}_m < \infty$. Hence on setting $r = n-i+1$ and $s = n-j+1$ we have that

$$0 \le \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\gamma_0(j-i) \le \mathcal{C}_m \sum_{r=1}^{n}\sum_{s=1}^{n} (rs)^{-(d+1)}|\gamma_0(r-s)|\,. \tag{A.6}$$

But $|\gamma_0(\tau)| \le \mathcal{C}_\varrho\gamma_0(0)|\tau|^{2d_0-1}$, $\mathcal{C}_\varrho < \infty$, for all $\tau \neq 0$, and

$$\begin{aligned}
\sum_{r=1}^{n}\sum_{s=1}^{n} (rs)^{-(d+1)}|\gamma_0(r-s)| &\le n^{-2(d+1)}\gamma_0(0)(n + 2\mathcal{C}_\varrho \sum_{k=1}^{n-1}(n-k)k^{2d_0-1}) \\
&\le n^{-(2d+1)}\gamma_0(0)(1 + 2\mathcal{C}_\varrho \sum_{k=1}^{n-1} k^{2d_0-1}) \\
&\sim \frac{\gamma_0(0)}{n^{(2d+1)}} \times \begin{cases} 1 + 2\mathcal{C}_\varrho\zeta(1-2d_0)\,, & d_0 < 0; \\ 1 + 2\mathcal{C}_\varrho \log n\,, & d_0 = 0; \\ 1 + 2\mathcal{C}_\varrho n^{2d_0}/2d_0\,, & d_0 > 0. \end{cases}
\end{aligned}$$

It follows that for all $d > -1/2$

$$E_0[\mathbf{Y}^T\mathbf{M}\mathbf{Y}] \le \frac{\mathcal{C}_m\gamma_0(0)}{n^{1-2(d_0-d)}} \times \begin{cases} 1 + 2\mathcal{C}_\varrho\zeta(1-2d_0)/n^{2d_0}\,, & d_0 < 0; \\ 1 + 2\mathcal{C}_\varrho \log n\,, & d_0 = 0; \\ 1 + \mathcal{C}_\varrho/d_0\,, & d_0 > 0; \end{cases}$$

and we can therefore conclude that

$$Pr\left(n^{-1}\mathbf{Y}^T\mathbf{M}\mathbf{Y} > \epsilon\right) < O(n^{2(d_0-d)-2})\epsilon$$

for all $\epsilon > 0$ by Markov's inequality. Since ϵ is arbitrary it follows that when $d > -1/2$ and $d_0 - d < 1/2$ the almost sure limit of $n^{-1}\mathbf{Y}^T\mathbf{M}\mathbf{Y}$ is zero, by the Borell-Cantelli lemma, giving the desired result.

A.4 Proof of Theorem 2:

First note that

$$Q_N(\boldsymbol{\eta}) = \left\{\frac{\sigma_0^2\Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\right\} K_N(\boldsymbol{\eta}) \tag{A.7}$$

by the same argument that gives (17). Now let $\Delta C_N(z) = \sum_{j=N+1}^{\infty} c_j z^j = C(z) - C_N(z)$. Then

$$\begin{aligned}
|C(e^{i\lambda})|^2 = |C_N(e^{i\lambda})|^2 &+ C_N(e^{i\lambda})\Delta C_N(e^{-i\lambda}) \\
&+ \Delta C_N(e^{i\lambda})C_N(e^{-i\lambda}) + |\Delta C_N(e^{i\lambda})|^2
\end{aligned}$$

and the remainder term can be decomposed as $R_N = R_{1N} + R_{2N}$ where

$$R_{1N} = \left(\frac{\sigma_{\varepsilon 0}^2}{2\pi}\right) \int_0^\pi |\Delta C_N(e^{i\lambda})|^2 |2\sin(\lambda/2)|^{-2(d_0-d)} d\lambda \tag{A.8}$$

and

$$R_{2N} = \left(\frac{\sigma_{\varepsilon 0}^2}{2\pi}\right) \int_{-\pi}^\pi \Delta C_N(e^{i\lambda}) C_N(e^{-i\lambda}) |2\sin(\lambda/2)|^{-2(d_0-d)} d\lambda\,. \tag{A.9}$$

The first integral in (A.8) equals

$$\left\{\frac{\sigma_0^2 \Gamma(1-2(d_0-d))}{2\Gamma^2(1-(d_0-d))}\right\} \left(\sum_{j=N+1}^\infty c_j^2 + 2\sum_{k=N+1}^\infty \sum_{j=k+1}^\infty c_j c_k \rho(j-k)\right).$$

Because $B(z) \neq 0$, $|z| \leq 1$, it follows that $|c_j| < \mathcal{C}\zeta^j$, $j = 1, 2, \ldots$, for some $\mathcal{C} < \infty$ and $\zeta \in (0,1)$, and hence that

$$\sum_{j=N+1}^\infty c_j^2 < \zeta^{2(N+1)} \frac{\mathcal{C}^2}{(1-\zeta^2)}\,.$$

Furthermore, since $0 < d, d_0 < 0.5$ it follows that $|d_0 - d| < 0.5$ and Sterling's approximation can therefore be used to show that $|\rho(h)| < \mathcal{C}'^{2(d_0-d)-1}$, $h = 1, 2, \ldots$, for some $\mathcal{C}' < \infty$. This implies that

$$\begin{aligned} \left|\sum_{k=N+1}^\infty \sum_{j=k+1}^\infty c_j c_k \rho(j-k)\right| &< \sum_{r=0}^\infty \sum_{s=r+1}^\infty \mathcal{C}^2 \mathcal{C}'^{2(N+1)} \zeta^r \zeta^s (s-r)^{2(d_0-d)-1} \\ &< \zeta^{2(N+1)} \frac{\mathcal{C}^2 \mathcal{C}'}{(1-\zeta)^2}\,. \end{aligned}$$

Thus we can conclude that $R_{1N} < const.\, \zeta^{2(N+1)}$ where $0 < \zeta < 1$. Applying the Cauchy-Schwarz inequality to the second integral in (A.9) enables us to bound $|R_{2N}|$ by $2(\sigma_{\varepsilon 0}/\sigma)\sqrt{I_N \cdot R_{1N}}$. It therefore follows from the preceding analysis that $|R_{2N}| < const.\, \zeta^{(N+1)}$. Since $|R_N| \leq R_{1N} + |R_{2N}|$ and $(N+1)/\exp(-(N+1)\log\zeta) \to 0$ as $N \to \infty$ it follows that $R_N = o(N^{-1})$, as stated. The gradient vector of $Q(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ is

$$\frac{\partial Q(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \left(\frac{\sigma_0^2}{2\pi}\right) \int_{-\pi}^\pi \frac{C(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}} \frac{\partial}{\partial \boldsymbol{\eta}} \{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\} d\lambda$$

and substituting $C(z) = C_N(z) + \Delta C_N(z)$ gives $\partial Q(\boldsymbol{\psi})/\partial \eta_j = \partial Q_N(\boldsymbol{\eta})/\partial \eta_j + R_{3N} + R_{4N}$ for the typical j'th element where

$$R_{3N} = \left(\frac{\sigma_0^2}{2\pi}\right) \int_{-\pi}^\pi \frac{C_N(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}} \frac{\partial}{\partial \eta_j} \{\frac{\Delta C_N(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\} d\lambda$$

and

$$R_{4N} = \left(\frac{\sigma_0^2}{2\pi}\right) \int_{-\pi}^\pi \frac{\Delta C_N(e^{i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}} \frac{\partial}{\partial \eta_j} \{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\} d\lambda\,.$$

The Cauchy-Schwarz inequality now yields the inequalities

$$|R_{3N}|^2 \leq R_{1N} \left(\frac{\sigma_0^2}{2\pi}\right) \int_{-\pi}^\pi \frac{|C_N(e^{i\lambda})|^2}{|2\sin(\lambda/2)|^{2(d_0-d)}} \left|\frac{\partial}{\partial \eta_j} \{\log \frac{\Delta C_N(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}\right|^2 d\lambda$$

and

$$|R_{4N}|^2 \leq R_{1N} \left(\frac{\sigma_0^2}{2\pi}\right) \int_{-\pi}^\pi \left|\frac{\partial}{\partial \eta_j} \{\frac{C(e^{-i\lambda})}{|2\sin(\lambda/2)|^{(d_0-d)}}\}\right|^2 d\lambda\,,$$

from which we can infer that $|R_{3N} + R_{4N}| \leq const.\, \zeta^{(N+1)} = o(N^{-1})$, thus completing the proof.

A.5 Proof of Theorem 3:

To establish (26) we will first show that for the Whittle estimator we have $\mathcal{C}_n(\boldsymbol{\eta})\partial Q_n^{(2)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta} = \partial Q_n^{(1)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta} + o(n^{-1/2})$ where $\mathcal{C}_n(\boldsymbol{\eta}) = Q_n^{(1)}(\boldsymbol{\eta})$. For the TML and CSS estimators we will then show that $2R_n\partial Q_n^{(3)}(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}$ and $R_n\partial Q_n^{(2)}(\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}$ converge in distribution, and that $n^{1/2}\partial\{Q_n^{(4)}(\boldsymbol{\eta}) - Q_n^{(2)}(\boldsymbol{\eta})\}/\partial\boldsymbol{\eta} = o_p(1)$, respectively. For the Whittle estimator we have

$$\frac{\partial Q_n^{(2)}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} = \frac{1}{Q_n^{(1)}(\boldsymbol{\eta})}\frac{\partial Q_n^{(1)}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} + \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial\log\left[f_1(\boldsymbol{\eta},\lambda_j)\right]}{\partial\boldsymbol{\eta}}\,.$$

By Proposition 1 $Q_n^{(1)}(\boldsymbol{\eta})$ converges to a positive constant, and following the development in Chen and Deo's proof of their Lemma 4 (see Chen and Deo, 2006, p. 270), the deterministic function

$$\frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial\log\left[f_1(\boldsymbol{\eta},\lambda_j)\right]}{\partial\boldsymbol{\eta}} = O(n^{-1}\log^2 n) = o(n^{-1/2})\,.$$

It follows that $n^{1/2}\left(Q_n^{(1)}(\boldsymbol{\eta})\,\partial Q_n^{(2)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta} - \partial Q_n^{(1)}(\boldsymbol{\eta})\}/\partial\boldsymbol{\eta}\right) = o(1)$ almost surely. The asymptotic equivalence of the two estimators now follows; in Case 1 because $n^{1-2d^*}/n^{1/2}\log n \to 0$ as $n\to\infty$ when $d^* > 0.25$, in Case 2 because $n^{1/2}[\overline{\Lambda}_{dd}]^{-1/2} \propto \left(n/\log^3 n\right)^{1/2}$ when $d^* = 0.25$ by Lemma 10 of Chen and Deo (2006) and, trivially, $1/\log^{3/2} n \to 0$ as $n\to\infty$, and directly in Case 3 when $d^* < 0.25$. For $R_n\partial\{2Q_n^{(3)}(\boldsymbol{\eta}_1) - Q_n^{(2)}(\boldsymbol{\eta}_1)\}/\partial\boldsymbol{\eta}$ we begin by noting that by Theorem 3 of Lieberman and Phillips (2004) (See also Dahlhaus, 1989, Theorem 5.1), and definition of the Riemann-Stieltjes integral,

$$\frac{1}{n}\frac{\partial\log|\boldsymbol{\Sigma}_\eta|}{\partial\boldsymbol{\eta}} = \frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial\boldsymbol{\Sigma}_\eta}{\partial\boldsymbol{\eta}} \sim \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial\log\left[f_1(\boldsymbol{\eta},\lambda_j)\right]}{\partial\boldsymbol{\eta}}\,.$$

Our task therefore reduces to a consideration of the properties of

$$\frac{1}{n}\frac{\partial\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}}{\partial\boldsymbol{\eta}} - \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}I(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}}\,,$$

which we rewrite as $\mathbf{a}-\mathbf{b}$ where

$$\mathbf{a} = \frac{1}{n}\frac{\partial\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}}{\partial\boldsymbol{\eta}} - \frac{1}{n}\mathrm{tr}\frac{\partial\boldsymbol{\Sigma}_\eta^{-1}}{\partial\boldsymbol{\eta}}\boldsymbol{\Sigma}_0$$

and

$$\mathbf{b} = \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{I(\lambda_j)}{f_0(\lambda_j)} - 1\right)f_0(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}}$$

recognizing, via Theorem 3 of Lieberman and Phillips (2004) once again, that

$$\begin{aligned}
E_0\left(\frac{1}{n}\frac{\partial\mathbf{Y}^T\boldsymbol{\Sigma}_\eta^{-1}\mathbf{Y}}{\partial\boldsymbol{\eta}}\right) &= \frac{1}{n}\mathrm{tr}\frac{\partial\boldsymbol{\Sigma}_\eta^{-1}}{\partial\boldsymbol{\eta}}\boldsymbol{\Sigma}_0 \\
&= -\frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial\boldsymbol{\Sigma}_\eta}{\partial\boldsymbol{\eta}}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0 \\
&\sim \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}f_0(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}}\,.
\end{aligned}$$

Using expression (A.10) below we can therefore deduce that

$$E_0(\mathbf{a}-\mathbf{b}) = \begin{cases} O(n^{2d^*-1}\log n), & 0 < d^* < 0.5\,; \\ O(n^{-1}\log^3 n), & 0.5 < d^* \leq 0\,. \end{cases}$$

From the binomial expansion of $(a-b)^r$, $r \geq 2$, it follows that the higher order cumulants will converge to zero if the corresponding cumulants of $a = \boldsymbol{\lambda}^T\mathbf{a}$ and $b = \boldsymbol{\lambda}^T\mathbf{b}$ are asymptotically equal (modulo a constant multiple) for every fixed vector $\boldsymbol{\lambda} \neq \mathbf{0}$. The desired result then follows, implicitly invoking the Cramér-Wold device, since the cumulants are convergence determining for the limiting distributions in Theorem 3. We will show that a and b asymptotically share the same cumulants in the special case where $\boldsymbol{\lambda}^T = (1, 0, \ldots, 0)$. This corresponds to considering the asymptotic distribution of the estimate of d and demonstrates the detailed particulars required to deal with the two critical cases involving convergence rates less than $n^{1/2}$. Denoting the rth cumulant of a by $\kappa_0^r(a)$, we obtain for $r \geq 2$

$$\begin{aligned} \kappa_0^r(a) =& n^{-r}(r-1)!2^{r-1}\mathrm{tr}\left\{\frac{\partial \boldsymbol{\Sigma}_\eta^{-1}}{\partial d}\boldsymbol{\Sigma}_0\right\}^r \\ \sim& n^{-(r-1)}(r-1)!2^{r-1}\frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\right)^r\left(\frac{\partial\{\log f_1(\boldsymbol{\eta},\lambda_j)\}}{\partial d}\right)^r, \end{aligned}$$

using Theorem 3 of Lieberman and Phillips (2004) once more. For b, let

$$\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} y_t \exp(-i\lambda t) = \xi_c(\lambda) - i\xi_s(\lambda)$$

and set $\mathbf{X}^T = (\xi_c(\lambda_1), \xi_s(\lambda_1), \ldots, \xi_c(\lambda_{\lfloor n/2\rfloor}), \xi_s(\lambda_{\lfloor n/2\rfloor})\mathbf{F}_0^{-1/2}$ where

$$\mathbf{F}_0 = \mathrm{diag}(f_0(\lambda_1), f_0(\lambda_1), \ldots, f_0(\lambda_{\lfloor n/2\rfloor}), f_0(\lambda_{\lfloor n/2\rfloor}))\,.$$

Then $\mathbf{X}^T$ is Gaussian with zero mean and covariance $\mathbf{I}+\boldsymbol{\Delta}$ where $\triangle_{jk} = O(j^{-d_0}k^{d_0-1}\log k)$ for $1 \leq j \leq k \leq \lfloor n/2\rfloor$, (see Moulines and Soulier ,1999, Lemma 4). Moreover,

$$\frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor} I(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial d} = \frac{2}{n}\mathbf{X}^T\mathbf{F}_0\mathbf{D}_1\mathbf{X}$$

where $\mathbf{D}_1 = \partial\mathbf{F}_1^{-1}/\partial d$,

$$\mathbf{F}_1 = \mathrm{diag}(f_1(\boldsymbol{\eta},\lambda_1), f_1(\boldsymbol{\eta},\lambda_1)\ldots, f_1(\boldsymbol{\eta},\lambda_{\lfloor n/2\rfloor}), f_1(\boldsymbol{\eta},\lambda_{\lfloor n/2\rfloor}))\,,$$

from which it follows that

$$\begin{aligned} \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{E_0(I(\lambda_j))}{f_0(\lambda_j)}-1\right)f_0(\lambda_j)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial d} &= \frac{2}{n}\mathrm{tr}\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta} \qquad \text{(A.10)} \\ &= O\left(n^{2d^*-1}\sum_{j=1}^{\lfloor n/2\rfloor} j^{-(1+2d^*)}\log^2 j\right) \\ &= \begin{cases} O(n^{2d^*-1}\log n), & 0 < d^* < 0.5\,; \\ O(n^{-1}\log^3 n), & -0.5 < d^* \leq 0\,, \end{cases} \end{aligned}$$

since

$$\frac{\partial f_1(\boldsymbol{\eta},\lambda)^{-1}}{\partial d} = -2\frac{\log 2|\sin\lambda/2|}{f_1(\boldsymbol{\eta},\lambda)} = O(\lambda^{2d_1}\log\lambda)\,,$$

cf. Lemma 4 of Chen and Deo (2006). For $r \geq 2$ we have

$$\kappa_0^r(b) = 2^r n^{-r}(r-1)!2^{r-1}\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1(\mathbf{I}+\boldsymbol{\Delta})\right\}^r$$

and the expansion $\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1(\mathbf{I}+\boldsymbol{\Delta})\right\}^r = \sum_{j=0}^r \binom{r}{j}\mathrm{tr}\{\mathbf{F}_0\mathbf{D}_1\}^{r-j}\{\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta}\}^j$ yields the result that

$$\begin{aligned}\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1(\mathbf{I}+\boldsymbol{\Delta})\right\}^r = \mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r + &\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r\boldsymbol{\Delta} \\ &+ O\left(\sum_{j=2}^r \binom{r}{j}\mathrm{tr}\{\mathbf{F}_0\mathbf{D}_1\}^{r-j}\{\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta}\}^j\right). \qquad \text{(A.11)}\end{aligned}$$

Evaluating the terms on the right hand side of (A.11) gives

$$\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r = 2\sum_{j=1}^{\lfloor n/2\rfloor}\left(\frac{f_0(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\right)^r\left(\frac{\partial\{\log f_1(\boldsymbol{\eta},\lambda_j)\}}{\partial d}\right)^r,$$

$$\begin{aligned}\mathrm{tr}\left\{\mathbf{F}_0\mathbf{D}_1\right\}^r\boldsymbol{\Delta} &= O\left(n^{2rd^*}\sum_{j=1}^{\lfloor n/2\rfloor} j^{-(1+2rd^*)}\log^{(r+1)} j\right) \\ &= \begin{cases} O(n^{2rd^*}\log n), & 0 < d^* < 0.5\,; \\ O(\log^{(r+2)} n), & -0.5 < d^* \leq 0\,,\end{cases}\end{aligned}$$

and, similarly

$$\mathrm{tr}\{\mathbf{F}_0\mathbf{D}_1\}^{r-j}\{\mathbf{F}_0\mathbf{D}_1\boldsymbol{\Delta}\}^j = \begin{cases} O(n^{2rd^*}\log^{(j-1+2d_0)} n), & 0 < d^* < 0.5\,; \\ O(\log^{(r+2(1+d_0))} n), & -0.5 < d^* \leq 0\,,\end{cases}$$

for $j = 2,\ldots,r$. It follows that

$$\frac{\kappa_0^r(2a)-\kappa_0^r(b)}{(r-1)!2^{r-1}} = \begin{cases} O(n^{r(2d^*-1)}\log n) + \sum_{j=2}^r\binom{r}{j}O(n^{r(2d^*-1)}\log^{(j-1+2d_0)} n), & 0 < d^* < 0.5\,; \\ O(n^{-r}\log^{(r+2)} n) + O(n^{-r}\log^{(r+2(1+d_0))} n), & -0.5 < d^* \leq 0\,,\end{cases}$$

leading to the desired result, namely that $R_n\partial\{2Q_n^{(3)}(\boldsymbol{\eta}_1) - Q_n^{(2)}(\boldsymbol{\eta}_1)\}/\partial d = o_p(1)$ where $R_n = n^{1-2d^*}/\log n$ when $d^* > 0.25$, Case 1, $R_n = (n/\log^3 n)^{1/2}$ when $d^* = 0.25$, Case 2, and $R_n = n^{1/2}$ when $d^* < 0.25$, Case 3. The corresponding results for arbitrary $\boldsymbol{\lambda} \neq \mathbf{0}$ can be obtained by reexpressing the linear combinations as $a = \boldsymbol{\lambda}^T\mathbf{a} = \partial Q_n^{(a)} - E_0(\partial Q_n^{(a)})$ and $b = \boldsymbol{\lambda}^T\mathbf{b} = \partial Q_n^{(b)} - E_0(\partial Q_n^{(b)})$ where the quadratic forms are given by

$$\partial Q_n^{(a)} = \frac{1}{n}(\mathbf{Y}^T\otimes(1,\ldots,1))\left[\langle\frac{\partial}{\partial\boldsymbol{\eta}}\{\boldsymbol{\Sigma}_\eta^{-1}\}\rangle\otimes\langle\boldsymbol{\lambda}\rangle\right](\mathbf{Y}\otimes(1,\ldots,1)^T),$$

where

$$\left[\langle\frac{\partial}{\partial\boldsymbol{\eta}}\{\boldsymbol{\Sigma}_\eta^{-1}\}\rangle\otimes\langle\boldsymbol{\lambda}\rangle\right] = \mathrm{diag}(\frac{\partial}{\partial\eta_1}\{\boldsymbol{\Sigma}_\eta^{-1}\},\ldots,\frac{\partial}{\partial\eta_{l+1}}\{\boldsymbol{\Sigma}_\eta^{-1}\})\otimes\mathrm{diag}(\lambda_1,\ldots,\lambda_{l+1}),$$

and

$$\partial Q_n^{(b)} = \frac{1}{n}(\mathbf{X}^T\otimes(1,\ldots,1))\left[\langle\frac{\partial}{\partial\boldsymbol{\eta}}\{\mathbf{F}_0\mathbf{F}_1^{-1}\}\rangle\otimes\langle\boldsymbol{\lambda}\rangle\right](\mathbf{X}\otimes(1,\ldots,1)^T).$$

The cumulants of a and b of order $r \geq 2$ can then be evaluated in the same manner as described above for the special case $\boldsymbol{\lambda}^T = (1,0,\ldots,0)$, the remaining details involve only more

complex notational and bookkeeping conventions. For the difference between $\partial Q_n^{(4)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta}$ and $\partial Q_n^{(2)}(\boldsymbol{\eta})/\partial\boldsymbol{\eta}$ we have

$$\frac{\partial\{Q_n^{(4)}(\boldsymbol{\eta}_1)\}}{\partial\boldsymbol{\eta}} = \frac{\partial}{\partial\boldsymbol{\eta}}\{\frac{\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}}{n}\} - \frac{\partial}{\partial\boldsymbol{\eta}}\{\frac{\mathbf{Y}^T\mathbf{M}_\eta\mathbf{Y}}{n}\},$$

and by Lemma A.1 it follows that

$$\frac{\partial}{\partial\boldsymbol{\eta}}\{\frac{\mathbf{Y}^T\mathbf{A}_\eta\mathbf{Y}}{n}\} - \frac{2}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{I(\lambda_j)}{f_1(\boldsymbol{\eta},\lambda_j)}\frac{\partial\{\log f_1(\boldsymbol{\eta},\lambda_j)\}}{\partial\boldsymbol{\eta}} = o_p(n^{-1/2}).$$

Now let $\dot\eta = (\eta_1,\ldots,\dot\eta_j,\ldots,\eta_{l+1)})^T$ and set

$$\nabla\mathbf{M}(\dot\eta_j) = \begin{cases} \frac{\mathbf{M}_{\dot\eta}-\mathbf{M}_\eta}{\dot\eta_j-\eta_j}, & \dot\eta_j \neq \eta_j\,; \\ \frac{\partial\{\mathbf{M}_\eta\}}{\partial\eta_j}, & \dot\eta_j = \eta_j\,. \end{cases}$$

Then $\lim_{\dot\eta_j\to\eta_j}\nabla\mathbf{M}(\dot\eta_j) = \nabla\mathbf{M}(\eta_j)$ and for all $\dot\eta_j \neq \eta_j$ we can employ an argument that parallels that following (A.6) to deduce that

$$Pr\left(n^{-1/2}\left|\mathbf{Y}^T\nabla\mathbf{M}(\dot\eta_j)\mathbf{Y}\right| > \epsilon\right) = O(n^{-(3+2d)/2})$$

for all $\epsilon > 0$, and hence that

$$n^{-1/2}\frac{\partial\{\mathbf{Y}^T\mathbf{M}_\eta\mathbf{Y}\}}{\partial\eta_j} = \lim_{\dot\eta_j\to\eta_j}\frac{\mathbf{Y}^T\nabla\mathbf{M}(\dot\eta_j)\mathbf{Y}}{n^{1/2}} = o_p(1).$$

This establishes that $n^{1/2}\partial\{Q_n^{(4)}(\boldsymbol{\eta}) - Q_n^{(2)}(\boldsymbol{\eta})\}/\partial\boldsymbol{\eta} = o_p(1)$, and the asymptotic equivalence stated in (26) now follows, because $n^{1-2d^*}/n^{1/2}\log n \to 0$ as $n\to\infty$ in Case 1, in Case 2 because $1/\log^{3/2} n \to 0$ as $n\to\infty$, and directly in Case 3. The preceding derivations, in conjunction with (25), imply that for the Whittle estimator $R_n(\widehat{\boldsymbol{\eta}}_1^{(2)} - \widehat{\boldsymbol{\eta}}_1^{(1)}) \to^D 0$, and that for the TML and CSS estimators $R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(2)}) \to^D 0$ for $i = 3$ and 4, for all three values of R_n. The asymptotic equivalence of all four estimators now follows since an immediate corollary is that $R_n(\widehat{\boldsymbol{\eta}}_1^{(i)} - \widehat{\boldsymbol{\eta}}_1^{(j)}) \to^D 0$, $i,j = 1,2,3$ and 4, for all three values of R_n.

B Evaluation of Bias Correction Term

For the FML estimator we have

$$\begin{aligned} E_0\left(\frac{\partial Q_n^{(1)}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right) &= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}E_0(I(\lambda_j))\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}} \\ &= \frac{2\pi}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\sum_{|k|<n}\left(1-\frac{|k|}{n}\right)\gamma_0(k)\exp(ik\lambda_j)\right)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}}, \end{aligned}$$

where $\gamma_0(k)$ denotes the autocovariance at lag k of the TDGP (see, for example, Brockwell and Davis, 1991, Proposition 10.3.1). Similarly, for the Whittle estimator we have

$$\begin{aligned} E_0\left(\frac{\partial Q_n^{(2)}(\sigma_\varepsilon^2,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right) = \frac{4}{n}\sum_{j=1}^{\lfloor n/2\rfloor}\frac{\partial\log f_1(\boldsymbol{\eta}_1,\lambda_j)}{\partial\boldsymbol{\eta}} \\ + \frac{8\pi}{\sigma^2 n}\sum_{j=1}^{\lfloor n/2\rfloor}\left(\sum_{|k|<n}\left(1-\frac{|k|}{n}\right)\gamma_0(k)\exp(ik\lambda_j)\right)\frac{\partial f_1(\boldsymbol{\eta},\lambda_j)^{-1}}{\partial\boldsymbol{\eta}}. \end{aligned}$$

Differentiating the TML criterion function with respect to $\boldsymbol{\eta}$ gives

$$\frac{\partial Q_n^{(3)}(\sigma^2,\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial \boldsymbol{\Sigma}_\eta}{\partial \boldsymbol{\eta}} + \frac{1}{n\sigma^2}\mathbf{Y}^T\frac{\partial \boldsymbol{\Sigma}_\eta^{-1}}{\partial \boldsymbol{\eta}}\mathbf{Y}\,,$$

which has expectation

$$E_0\left(\frac{\partial Q_n^{(3)}(\sigma^2,\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right) = \frac{1}{n}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial \boldsymbol{\Sigma}_\eta}{\partial \boldsymbol{\eta}} - \frac{1}{n\sigma^2}\mathrm{tr}\boldsymbol{\Sigma}_\eta^{-1}\frac{\partial \boldsymbol{\Sigma}_\eta}{\partial \boldsymbol{\eta}}\boldsymbol{\Sigma}_\eta^{-1}\boldsymbol{\Sigma}_0\,,$$

where $\boldsymbol{\Sigma}_0 = [\gamma_0(|i-j|)]$ and $\sigma^2\boldsymbol{\Sigma}_\eta = [\gamma_1(|i-j|)]$, $i,j = 1,2,...,n$. The criterion function for the CSS estimator in (10) can be re-written as

$$Q_n^{(4)}(\boldsymbol{\eta}) = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{i=0}^{t-1}\tau_i y_{t-i}\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\tau_i\tau_j y_{t-i}y_{t-j}\,,$$

where τ_i is as defined in (12). The gradient of $Q_n^{(4)}(\boldsymbol{\eta})$, recalling that $\tau_i = \tau_i(\boldsymbol{\eta})$, is thus given by

$$\frac{\partial Q_n^{(4)}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\left(\tau_i\frac{\partial \tau_j}{\partial \boldsymbol{\eta}} + \tau_j\frac{\partial \tau_i}{\partial \boldsymbol{\eta}}\right)y_{t-i}y_{t-j}\,,$$

and the expected value of the gradient is

$$E_0\left(\frac{\partial Q_n^{(4)}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right) = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\left(\tau_i\frac{\partial \tau_j}{\partial \boldsymbol{\eta}} + \tau_j\frac{\partial \tau_i}{\partial \boldsymbol{\eta}}\right)\gamma_0(i-j)\,.$$